DOCUMENT RESUME

ED 211 593 TM 820 027

AUTHOR	Choppin, Bruce H.; And Others
TITLE	Test Use Project. Annual Report.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	National Inst. of Education (ED), Washington, D.C.
PUB DATE	Nov 81
GRANT	NIE-G-80-0112
NOTE	142p.
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Elementary Secondary Education; Grade 4; Grade 6; Grade 10; Language Arts; Mathematics Achievement; *National Surveys; Reading Achievement; *Teacher Attitudes; Testing Problems; *Test Use

ABSTRACT

One of the two major phases of the Test Use Project of the Center for the Study of Evalaution (CSE) is discussed; that is, the collection and analyses of survey data from a national sample of teachers and principals representing the targeted grades/schools. Some historical background influencing that phase of the project is provided. The findings emanating from CSE's 1978 small-scale study of testing--which in some ways was the primary inception for the study--are described. The findings of fieldwork preceding the national survey are discussed and principal findings of the survey are presented. The survey included questions about the teacher's professional background, classroom characteristics, use of resources, district assistance, district training and college courses, district uses of assessment information, district reporting of test results, teacher attitude toward tests and test-related issues, and teacher uses of assessment results. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Center for the Study of Evaluation

UCLA Graduate School of Education
Los Angeles, California 90024

ED211593

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

CSE
CSECSE
CSECSECSE
CSECSE
CSE

TM 820 027

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. C. Beer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

DELIVERABLE - November 1981

TEST USE PROJECT
Annual Report

Bruce H. Choppin, Project Co-Director
Donald W. Dorr-Bremme, Project Co-Director
James Burry, Senior Research Associate

Grant Number
NIE-G-80-0112
P-2

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California . Los Angeles

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Table of Contents

Page

Introduction -- 1

Project History --- 4

- Planning Activities
- The Research Model Guiding the Test Use Project

Findings from the 1978 Study ------------------------------ 14

- The Volume of Testing Occurring in Schools
- The Extent to Which Teachers Use Test Results
- Teachers' Knowledge of and Attitude Toward Tests
- Factors Influencing the Use of Test Results
- Summary

The Exploratory Fieldwork --------------------------------- 28

- Intentions
- General Findings
- Range of Tests Administered
- Range of Reported Uses
- Patterns of Assessment Results Use
- Relationships Between Types of Tests and Categories of Use
- District Narratives
- District One
- District Two
- District Three
- Test Uses/Issues in District One
- District Summary
- Test Uses/Issues in District Two
- District Summary
- Test Uses/Issues in District Three
- District Summary

The National Survey --------------------------------------- 65

- Sampling Methodology
- The Initial Plan
- Revised Sampling Design
- First Stage (Selection of Districts)
- Stratification
- Sampling Frame
- Selection Procedure
- Second Stage (Selection of Schools)
- Third Stage (Selection of Teachers

Page

The Elementary School Teacher Sample ---------------------- 90

Teachers' Professional Background
Classroom Characteristics
Use of Resources
District Assistance
District Training/College Courses
District Uses of Assessment Information
District Reporting of Test Results
Teacher Attitude Toward Tests and Test-Related Issues
Teacher Uses of Assessment Results

The Secondary School Teacher Sample ---------------------- 107

Teachers' Professional Background
Classroom Characteristics
Use of Resources
District Assistance
District Training/College Courses
District Uses of Assessment Information
District Reporting of Test Results
Teacher Attitude Toward Testing and Test-Related Issues
Teacher Uses of Assessment Results

References -- 131

INTRODUCTION

CSE's Test Use Project has been gathering information bearing on a range of testing issues for students, teachers, administrators, researchers, and policy makers. It is clear that our schools do a great deal of student achievement testing, and some limited information has already been collected on certain practices affecting our students in some areas of the country. Until the CSE study, however, we have lacked information that is nationally representative and illustrative of the entire range of tests being administered, and yet which is sufficiently focused to be of use in test-based policy matters.

CSE has been concerned, first, that we have been lacking descriptive data reflecting the entire testing picture--the range of tests being administered, their associated users and consumers, and the range of students affected by particular kinds of tests. Second, we have also lacked the more inferential utilization data--the primary and secondary users of test information, the intended and actual uses of test information, variations in use across users and organizational settings, the kinds of decisions made on the basis of test information, the kinds of students thereby affected, and the attendant costs of the testing enterprise.

Since the inception of the Test Use Project in December 1979, we have been examining these kinds of issues in a broad framework which defines testing to include formal tests, both norm- and criterion-referenced; curriculum-embedded measures; district-, school, and teacher-developed tests; as well as the more informal measures such as teacher quizzes, observations, and other interactions with students. In short, our study has not aimed at any single kind of test, user, or student. But the study is also sharply focused in this broad framework, and examines some of the more troublesome aspects of testing: student achievement testing in language arts and mathematics; at selected grade levels where testing may critically affect large numbers of students and their teachers--fourth and sixth grades in elementary schools and tenth grade in high schools; with emphasis on first- and second-orders of test use (Baker, 1978). Finally, information on these matters has been primarily reported to us by teachers and principals--those who are closely involved in first- and second-order uses of tests.

The Test Use Project has been proceeding in two overlapping phases. Phase I, taking place between December 1979 and November 1981, has culminated in the collection and analyses of survey data from a national sample of teachers and principals representing the targeted grades/schools. During Phase II of the study, which began in February 1981 and will conclude in November 1982, the project will be conducting on-site studies in a small number of schools. The primary intention of this phase of the study is to identify the direct and indirect costs of testing, with the secondary intention of pursuing salient findings of the Phase I work and expanding the contextual base critical to interpretation of its survey data.

In our work thus far we have developed and refined the conceptual scheme informing our work; reviewed and reported on the relevant literature; conducted preliminary fieldwork in schools to pilot-test questions about testing with teachers and principals; drawn a nationally representative sample of teachers and principals from the target grade levels; pilot-tested and sent out questionnaires to the sample;

analyzed the data resulting from that sample; and planned the conceptual framework for our Phase II activities.

When all analyses of the Phase I survey data have been completed, we plan to begin dessemination of results to teachers, principals, and other administrators, planners and policy makers, researchers, and testing specialists. Dissemination will continue through the Phase II cost study which, when completed, will relate testing practices to a range of monetary, opportunity, and psychological costs. Our findings should have a bearing not only on testing practices and test-based decisions about individuals and groups of students, but also on test related policy making and school practices including test selection, development, and use, as well as teacher inservice in these areas.

Since this report discusses one of the Test Use Project's two major phases, we will provide some of the historical background influencing that phase of our work and which led up to this document. We will also describe some of the findings emanating from CSE's 1978 small-scale study of testing which in many ways was the primary inception for the present study; continue the findings in a discussion of the field-work which preceded the national survey, and present the principal findings of the national survey.

PROJECT HISTORY

Planning Activities

There is little doubt that testing in our schools has been increasing in response to federal and state program assessment requirements, accountability concerns, national and regional assessment needs, state mandated minimum competency requirements, and the expansion of curriculum embedded testing programs.

As with other highly visible activities, testing has become the subject of much controversy, and the legal and political systems have entered the debate. Testing proponents have argued that tests contribute to educational quality control, help in providing individualized instruction to students, and assist in improved educational decision making. Critics of testing, on the other hand, have described the arbitrary nature of current testing practice, have challenged tests for their biased properties, and questioned their appropriateness to contemporary education and its changing functions.

While there is some empirical information available about testing--six full standardized test batteries, on an average, are taken by a student during his or her school years, at least 90 percent of the LEAs in the country administer standardized tests to their students, over 40 states conduct a state assessment program and/or have adopted minimum competency legislation--we have been lacking nationally representative information about the nature of this vast amount of testing and how it is or is not being used in schools. CSE's Test Use Project has been collecting information to answer these questions over the past two years. That is, we have been attempting to document how much testing is going on in schools, what kinds of tests are administered and with what frequency, which of these tests are used or not used in the decisions affecting schooling, and at what costs. In addition, we have been examining the coordinate issue of the contextual factors which influence the administration of tests and the use of tests for instructional decision making.

The framework we devised to investigate these matters suggested that in order to understand current testing practices, we needed to have, for each type of test administered, information concerning its intended purposes, its characteristics, the context of administration, the actual use of its results, and the costs. This framework enabled us not only to describe the nature of testing, but also to explore relationships among the surrounding components listed above.

Within the framework described above, then, we have been gathering information on testing practices, test uses, and testing costs over the two phases of the project. In each of these phases, our research became progressively more focused, beginning with wide-ranging inquiry to provide a comprehensive view of relevant phenomena and perspectives, followed by the design of specific study questions and instruments to answer them, and finally the collection and analyses of data collected on the questions of interest.

Our planning activities, which were devoted to refining and focusing the questions of the study and the framework within which they were pursued, had several components: we re-examined our previous test use data collected in 1978; we conducted a literature search and review to examine research on testing and test use and to identify a range of salient policy issues; we consulted researchers, test specialists, school-level practitioners, and administrators on the policy issues and foci appropriate for the study; we conducted exploratory fieldwork to assess the relevance of the guiding framework as a tool to provide us with information on tests being administered, the kinds of purposes they serve, and the factors influencing their administration and/or use. These activities helped us to explicate the full range of tests and other assessment devices being administered and the kinds of factors that might influence test use.

Together, the information stemming from each of our planning stages suggested that consideration of three basic questions was necessary to provide a rational structure for delimiting the emphases of the national survey:

- What issues and questions about educational testing presently confront those who make educational testing policy?
- What information is presently available to inform those questions and issues? What kinds of information gaps remain? Of that information, which will be most useful?
- Of that useful information, which can be obtained at an appropriate level of specificity within the scope of the CSE project and its available resources?

These questions--concerned with issues in educational testing, status of current information on testing and test use, and definition of our research problem--structured our thinking about directions for the national servey. For example, the matter of current issues in testing raised a variety of questions of potential relevance to the survey and to policy makers in a variety of test-related areas. The matter of the emergence and proliferation of competency testing is one such question. With more than forty states operating minimum competency testing programs, some of which require the tests for promotion and graduation and others simply for checking students' basic educational progress, it seemed to us that decision makers at all levels need to know if and how these programs influence students' educational experiences and life chances, and if they do, to what extent and how equitably. Policy makers are also concerned with externally required testing for program evaluation, with its related concerns of accountability and compliance and the degree to which it may serve other educational purposes. Another matter concerns district continuum testing and its quality and effectiveness in improving local instruction. Teacher constructed tests and other assessment techniques comprise another important issue since teachers seem to spend significant amounts of their time administering their own tests and quizzes. What are the qualities of these tests that make them appear attractive and useful to teachers? And can these qualities be incorporated in other tests and testing procedures? Finally, the area of current issues also reflects matters of equity, testing costs/benefits, and potential misuse. Are certain kinds of students possibly being over-tested at the expense of receiving necessary instruction? Are students in general being tested too much? Are particular kinds of tests and testing programs worth the time, energy, and money invested in them? Which have the greatest benefits and under which conditions? What patterns and/or combinations of testing provide the highest payoff in terms of generating valid and reliable information at minimum costs?

The second of the three questions delimiting the study reflected the status of our current information on testing and test use. What information is currently available to inform those concerned with the kinds of decisions outlined above? Our literature review (see: Volume II, Test Use Report, November 1980) suggested that very little concrete information presently exists.

Opinion and argumentation dominate the published material on testing. Experts debate the merits of norm-referenced and criterion-referenced tests. Proponents and opponents of minimum competency testing argue their cases. The cultural and linguistic bias of certain tests are cited. Calls appear for the development and use of alternative assessment procedures and for more teacher training in testing. These and similar discussions have helped focus the issues that policy makers now face, but research to address those issues is in short supply.

Few national studies on testing and test use have been conducted. Those that have been center on teachers' attitudes toward and use of norm-referenced, standardized tests (e.g., Ebel, 1967; Goslin, 1967; Kirkland, 1971; Stetz & Beck, 1979). This emphasis recurs in most of the extant research on testing in particular states and localities (c.f., Angel, 1968; Boyd, et al, 1975; Hotvedt, 1978; Infantino, 1975; Rudman, 1978; Salmon-Cox, 1980), but contemporary concerns in the area of educational testing go well beyond standardized testing. Information is required on a wide range of tests and assessment practices.

Work by Yeh (1978), which is discussed later in this report, and others suggests that these concerns about gathering information on a wide variety of assessment techniques is valid. Our test use exploratory fieldwork, also discussed later in this report, further pointed up the relevance of these issues.

The appropriateness of giving attention to, and raising questions about, the full range of tests and other assessment procedures is also indicated in sociological studies of teaching (e.g., Lortie, 1975;

Kitsuse & Cicourel, 1963) and research on teachers' decision making. (c.f., Borko, 1978, Leiter, 1974; Mehan, 1974; Shavelson, 1977. See also Airasian, 1979). Yeh's (1978) work in test use provided us with a starting point for examining this range. The research of others, still in progress, will also begin to extend understanding of the current functions of different kinds of testing (e.g., Rudman, Kelly, Wanous, Mehrens, Clark, & Porter, 1980; Resnick & Resnick, 1978; Sproull & Zubrow, 1979; National Evaluation Systems, 1978). At present, however, little is known about the uses and impacts of teachers' observation- and interaction-based judgments or teacher-made assignments and tests. The same is true about the functions and influences of tests embedded in commercially produced curricula. And the information on minimum competency testing, testing for state and federal program evaluation, or the objective-based testing accompanying district-mandated skills continua is equally limited. Aside from the extant work on standardized testing, there are only a few, rather narrowly focused studies on one or another kind of test (e.g., Carducci-Bolchazy, 1978; Crew & Whitney, 1978).

The above overview of issues and available information leads to the third of the three delimiting questions of the study. That is, what should a national survey of testing practices and test use attempt to accomplish? Clearly, what policy makers and stakeholders in educational testing now need most urgently is basic, broadly based descriptive and inferential information. They need to know what is going on in schools nationwide with respect to assessment of student achievement. More specifically, they need to know what tests and other assessment practices matter and in what circumstances they matter in American public schools.

Matter, as used here, is construed in two ways--one quantative, the other qualitative. In the former sense, a type of test or assessment practice matters to the extent that it occurs widely in American schools. Thus, our national survey of testing was concerned with identifying the types of assessment instruments and practices which are administered generally and frequently. In the second, qualitative sense, a type of test or assessment practice matters to the extent that it has impact. Thus, our survey of testing had also to identify those types of tests and practices that significantly influence the lives of students and the activities of practitioners in schools.

Furthermore, our survey work would have to be attentive to two kinds of impacts or influences. Tests, of course, can impact on the lives of students (and their families) when scores from those tests count as major factors in decisions made about them; e.g., placement and grading decisions. Test scores can also influence students' lives and teachers' activities when they are used as criteria in evaluating and changing curriculum, allocating funds, or identifying teachers' professional needs for inservice training. But the Test Use Project's exploratory field research in three school districts, as will be seen later in this report, has also called attention to the impacts that tests can have by virtue of their very presence as required or recommended activities.

In summary, the initial research problem for the CSE national survey of testing was to document what types of assessment are extant in American elementary and high schools and to discern where particular types fall on the following "map":

Figure 1

An Initial Map of National Testing
and Test Use

Does a type of assessment:

OCCUR WIDELY?

HAVE IMPACT?	Most	Least
Most	Cell 1 Occurs widely Has great impact	Cell 2 Does not occur widely Has great impact
Least	Cell 3 Occurs widely Has little impact	Cell 4 Does not occur widely Has little impact

This very basic information is currently lacking, as earlier discussion has argued.

Discovering how types of tests and other assessment practices array on the above "map" would indicate (Cells 1 and 3) which types are now consuming significant amounts of administrators', teachers', and students' time and energy--and (in rough approximation) public dollars. Research toward this end was also intended to indicate which types of assessment instruments and procedures bear most heavily on students' educational experiences and life chances and upon the professional activities of practitioners in schools. Simultaneously, then, such a "map" of tests would enable those concerned with testing to identify the tests that matter most nationwide (Cell 1, as "mattering" has been defined here) and those that matter least (Cell 4)--and it would offer a rough-but-useful initial guide to those types of assessment activities for which costs may currently exceed benefits (Cell 3). Thus, the survey would attempt to facilitate sorting and prioritizing the range of issues and questions that confront those concerned with assessment of student achievement in its various forms, while providing a basic descriptive picture of assessment activities.

The second research problem for the national survey, as noted above, was to identify and describe in what circumstances particular types of assessment activity matter. The survey would, therefore, seek data so that the descriptive "map" in Figure 1 could be differentiated: so that patterns of test use and impact under different contextual conditions could be described. Types of tests may occur with a frequency and/or degree and type of impact that varies from urban schools to rural, from schools serving the economically advantaged to those serving the economically disadvantaged, from classrooms where teachers are more experienced to those where they are less so. Achieving a differentiated description of testing and test use of this kind can result in the identification of the factors that influence the use and impacts of particular kinds of tests and other assessment practices. Consequently, the description should afford an understanding of conditions that contribute to optimal use of particular kinds of tests and other assessment procedures.

Summary. Earlier decisions led the Test Use Project's national survey to focus upon:

- Achievement testing in language arts/reading and mathematics at the upper elementary and high school grade levels.
- Test uses of the first- and second-order (Baker, 1978), i.e., the uses of testing within schools.
- Information on the latter as reported by classroom teachers and principals.

More specifically, the national survey would gather basic descriptive information on:

- The frequency and distribution of a broad range of types of achievement testing and other achievement assessment practices.
- The impacts of those types of testing and practices, i.e.,
  - the particular purposes for which test scores and other assessment results are used and their importance in serving those purposes,
  - the influences those types of testing and assessment practices have by virtue of their very presence as required or recommended activities,
- The combinations of factors that influence the uses and impacts of particular types of achievement tests and other achievement assessment practices.

Patterns of responses to survey questions on the above issues (as seen later in this document) will help provide basic data on the benefits that accrue for students and practitioners from types of achievement tests and assessment practices. Negative impacts cited by respondents will help to formulate some of the costs of testing. In Phase II of the Test Use Project, when follow-up field research occurs, the monetary, opportunity, and psychological costs of testing will be the focus of inquiry. The project's exploratory fieldwork confirmed the wisdom of this earlier decision. Even when using interviews in the field, checking cost information was extremely difficult.

At this point in our planning activities, we began to approach the question of study design and data collection, and selection of a research model most appropriate to our endeavors.

The Research Model Guiding the Test Use Project

One end-result of our planning activities was the selection of the central questions which would guide the national survey. These questions were stated as follows:

1. With what frequency and distribution are particular types of test given in the upper elementary grades and high school?
2. In what ways do particular types of tests and testing impact upon schools and those within them;
   a. through their very presence, required or recommended?
   b. through utilization of their results?
3. What factors influence;
   a. where and how much particular types of testing are done?
   b. the ways types of tests, testing, and test score use impact upon schools and those within them?

As will be recalled from our 1980 Test Use report to the NIE, since our survey was intended to be both descriptive and analytic, we were concerned that our research meet the canons of descriptive validity. In selecting a theory of the nature of the phenomenon to be described, the researcher imposes a reality, consisting of a set of constructs and statements of relationship and function, on the phenomenon being described. This imposition of reality occurs as the researcher attempts to describe the activities and events that are taking place and the manner in which they are taking place. In fact, when he/she makes decisions about what to select for description and what to omit, the researcher imposes his/her own structure on what is "really" taking place. It is critical, then, that the researcher's constructs and assumed relationships bear a resemblance to those which participants in the phenomenon being described actually act upon. In brief, this means that the researcher's description of the phenomenon being studied should attempt to integrate both the researcher's and the participants' orientations and conceptualizations. Thus, our survey design proceeded from a conceptual standpoint that maintained contact with the orientations and purposes of educators in schools and at the same time addressed our own central policy and research concerns.

As discussed in our 1980 report to the NIE, the study's conceptualization involved two interlinking concepts: that of the teacher as practical reasoner and decision maker; and that of testing as an intervention.

As practical reasoners and decision makers, teachers orient their activities to the practical tasks they must accomplish in their everyday routines and do so in light of the practical contingencies and exigencies they face. Teachers, further, carry out their activities based on their understanding of a "world known in common and taken for granted" (Schutz, 1962). Our planning stage activities can be interpreted from this perspective. That teachers orient their efforts to the practical tasks that are central to their everyday lives and that they do orient to their practical exigencies was recurrently documented in data gathered during our planning activities. Further, teachers rely on consensually-supported and phenomenologically-based understanding to carry out their tasks.

In our 1980 annual report to the NIE, we cited evidence from our fieldwork demonstrating that:

- Teachers report their uses of test results as serving most heavily the functions that are at the core of teaching-as-practiced.

- The means of assessment that teachers report using most often and in the greatest variety of ways are those which facilitate the accomplishment of their practical activities under the exigencies they face.

- Teachers tend to use least those tests which fit least well with the practical demands of their everyday world.

- For given activities and decisions, teachers most often use the results of various types of assessment techniques collectively. Scores from one test or one type of test rarely serve alone as the basis for accomplishing a task.

- Teachers orient to the routine constitutive tasks and exigencies of teaching-as-practiced.

The second concept framing our project's survey inquiry was the concept of testing as an intervention. That is, whether required or recommended, tests, by virtue of their very presence in the teacher's world, can function as educational change agents. Our planning stage fieldwork suggested that tests can function as such in any one of three ways:

- Mandated tests can add new standards of accountability to the practical exigencies teachers must attend to in their everyday routines.

- Mandated tests can change the practical circumstances under which teaching and learning must be accomplished.

- Testing programs of particular kinds can facilitate accomplishment of the routine tasks of teaching-as-practiced by responding to the practical exigencies teachers face.

(Evidence supporting these findings, though previously discussed in our reports to the NIE, will once again be summarized in a subsequent section of this report. That section, in addition to summarizing the fieldwork, will also discuss CSE test use data preceding the fieldwork and the findings of the Test Use Project's national survey.)

In the foregoing discussion we have outlined the concepts of teacher as practical decision maker and testing as an intervention. These concepts served to orient the design of our national survey. The two concepts converge to provide a grounded theory of test use in schools and classrooms. It is a theory taking into account the purposes and constructs of participants in the phenomenon under examination, and it is a theory which permits issues to be addressed that are central to policymakers, stakeholders in the testing enterprise, and the community of researchers studying educational testing.

This theory of test use provided a heuristic for the informed selection of domains to be examined in our survey research and indicated some relationships for study among those domains. The domains were concerned with the following:

Federal/state/local testing requirements
Federal/state/local instructional programs
Organization of curriculum and instruction
Types of students served
Teachers' perceptions of the utility of tests and types of tests
Teachers' experience and training
District and local site leadership action
Types of tests given: purposes and frequency
Types of test score use
Impacts of tests

The CSE national survey findings reflect the above kinds of domains. The reader who wishes to examine these findings is invited to resume reading of this report at page 65, where the selection of the national sample, the development of instrumentation, and discussion of its findings begins. The reader who is interested in all of CSE's test use findings, those which led up to as well as those stemming from the national survey, should continue below in the section dealing with our 1978 study.

FINDINGS FROM THE 1978 STUDY

As mentioned previously, one of CSE's early activities in gathering information on teachers' test practices and test use began in 1978. Two-hundred sixty teachers participated in this small-scale study, representing 20 California elementary schools in urban, rural, and suburban areas and in low and higher socioeconomic communities. The results of these teachers' reports gave some preliminary answers to our questions of:

- The volume of testing occurring in schools.
- The extent to which teachers use test results.
- Teachers' knowledge of and attitude toward tests.
- Factors influencing the use of tests.

The Volume of Testing Occurring in Schools

All schools in the study administered yearly state assessment tests in grades one, two, three, and six, and all administered annual or semi-annual standardized norm-referenced test batteries to their students. A sizeable number were required, in addition, to give beginning and end of year assessments of a criterion-referenced or district continuum variety. As with all California schools, the schools in the study were involved in required minimum competency testing. While this listing of required tests is sizeable, it is not exhaustive. Other kinds of tests, teachers reported, constituted a much greater proportion of assessment activities in schools.

One of the survey questions addressed those tests administered routinely by classroom teachers in their normal instructional activities. Teachers reported more frequent testing in mathematics than in reading, but the frequency in both subject areas was substantial. A majority of the teachers reported giving weekly or daily mathematics tests, and eighty percent reported at least monthly mathematics testing. About one-third of the teachers administered weekly reading tests, and another third reported monthly reading tests. Testing in both subject areas was less frequent in the primary grades than in the upper elementary levels.

The Extent to Which Teachers Use Test Results

The survey investigated use from two perspectives: first, what sources of information were used to make particular instructional decisions; and second, what use was made of test results? The first perspective inquired about the use of tests relative to other sources of available information; the second asked more directly about the use of particular types of tests; but gave a more limited sense of relative value.

Teachers were asked what sources of information they used most frequently at the beginning of the school year to assess student skills. Fifty-eight percent reported that test results were most important for initial reading placement, and 66 percent reported using test results most often for initial mathematics placement.

While these findings implied that test results, and even those from required tests, provided important information early in the year, the picture changed as school got underway. When asked the sources of information they used to assess student progress throughout the year, teachers reported relying most heavily on interactions with students, informal assessments (e.g., oral quizzes, reading aloud), and the results of teacher developed tests. It seemed that the results of standardized tests were rarely used, and that curriculum embedded tests fared only slightly better.

Test results, then, seemed to provide the teacher with a quick and acceptable estimate of the ability of new students with whom the teacher was unfamiliar. However, once initial placements were made and teachers became more acquainted with their students, they stated that they were less likely to rely on standardized or curriculum tests as information about students' progress.

A similar picture emerged when teachers were asked more directly about how they use the results from their own tests and from required tests. Teachers indicated that they usually used the results of their own tests for several purposes: to make instructional decisions, to evaluate the effectiveness of their classroom program (e.g., teaching strategies, curriculum materials), and to provide information to others (e.g., parents, other teachers). Teachers also reported using tests to assign grades, but with somewhat less frequency.

In contrast, teachers stated that they used the results from required tests only infrequently for any of the above purposes. They seemed to use these tests relatively most often for reporting to parents or other staff and for evaluating the effectiveness of teaching methods and materials; but their reported frequencies were quite low. Required test results seemed to function for teachers as a standard of comparison, while teacher made tests reportedly were used more for instructional decision making.

Teachers' Knowledge of and Attitude Toward Tests

Most teachers reported some training, e.g., college courses and inservice sessions, in educational measurement. Thirty-nine percent reported two or more college courses related to educational testing, while 23 percent reported no college courses in this area. A majority also reported at least one inservice course in testing.

Despite this formal training in testing, however, teacher's responses about appropriate interpretations of common standardized test scores raised some questions about their levels of understanding. When presented with test results, only 50 percent of the teachers were able to interpret correctly percentile and grade equivalent scores--the two methods most frequently used for reporting standardized test scores.

Survey data about teachers' attitudes toward required testing were more consistent. Responses about how teachers evaluated the costs vs. the benefits of testing, their reactions to discontinuing required testing, and their opinions of what required tests measure portrayed a somewhat negative picture.

When asked to rate the amount of classroom time spent in required testing relative to the teacher and student benefits which accrued, teachers felt that a bit too much time was spent in testing. Similarly, they responded that teachers would react favorably to the discontinuation of testing, though again their responses were not extreme. Finally, teachers stated that they felt that their students' performance on required tests was influenced to some extent by the instruction they received, but they stated that they believed student's motivation, test-taking skills, unusual circumstances, and test quality were more important factors.

Factors Influencing the Use of Test Results

Two lines of inquiry suggested factors which influence the use of tests by teachers. First, teachers were asked what features they considered in formulating their own classroom testing programs. As stated in an earlier section of this report, we assumed that the more tests exemplified desired features, the greater the likelihood they would be used. A second avenue of inquiry was more empirical: what contextual variables were associated with more test use? e.g., teaching experience, classroom organization (team teaching vs. self-contained), grade level taught, and availability of classroom aides.

What test qualities were most important to teachers? Teachers reported that clear format, similarity to class material, and accurate prediction of achievement were the qualities they considered most important when choosing prepared tests. Similarly, when asked why they developed their own tests rather than using commercial tests, teachers cited suitability for their students and sensitivity to classroom instruction as critical reasons. They stated that lack of funds, of time to order tests, or of information about tests were unimportant influences. Intuitive validity appeared to be the essential feature for teachers: did the test match what was taught and did it provide a suitable context so that students could exhibit their skills? This criterion contrasts teachers' perceptions of required tests as being heavily influenced by students' test-taking skills and other extraneous influences.

What contextual factors seemed to be associated with test use? Certainly grade level appeared to exert a significant influence. Primary grade teachers reported administering fewer tests, that they were less likely to develop their own tests, and that they would react more positively to abolishing required tests than would upper elementary school teachers.

Years of teaching experience was also related to different patterns of test use. Younger teachers, i.e., those with less than eight years of teaching experience, appeared more skeptical of testing. These teachers, relative to their more experienced peers, appeared more likely to use their own tests and other less formal methods (e.g., work assignments, informal quizzes, students' place in the text) to assess student progress, and less likely to use the results of required, standardized, or curriculum embedded tests. They were also apparently less optimistic about the extent to which instruction influences students' performance on required tests, an opinion consistent with their reported behavior. Perhaps these younger teachers were influenced during their preservice training by relatively recent criterion-referenced testing methodologies, and were, therefore, more suspicious of published tests.

The presence of aides was also associated with more frequent use of assessment data. Teachers with classroom aides, compared with those without such assistance, reported greater use of curriculum-embedded tests and used student's place in their book and other informal assessments more often to monitor their students' progress. It may be that teachers felt that considerable record keeping was required to make good use of test data for instructional decision making, that a classroom aide would ease significantly the burden in this area, and thus might be instrumental to a teacher's use of test data. Further, the teachers might have been concerned about using test data for instructional decision making to identify and better meet individual needs. The availability of aides might make teachers feel that they then have more time to prescribe alternative settings for instruction, e.g., aides can give tutorial assistance, supervise small student groups, etc. Without instructional alternatives, however, teachers might feel less motivation to use test data, because they lack the resources to carry out more individualized prescriptions and/or needed remediations. Consistent with this hypothesis, teachers with aides appeared less likely to allow failing students to progress to the next instructional unit, and more likely to provide such students with remedial help, e.g., tutoring and additional practice.

Summary

The findings of CSE's 1978 study replicated those of other researchers: Teachers in the sample reported that they do not make much use of the many standardized tests they are required to administer. Furthermore, while they perhaps were not adamantly hostile in the face of required testing, their attitudes towards these tests, at best, were reserved. These attitudes may explain reported patterns of use--or non-use. Teachers' knowledge in testing, no doubt, was also a contributing factor.

The teachers reported that required standardized tests comprised only a small fraction of classroom assessment activities. Curriculum-embedded tests and particularly teacher-made tests were not only more prevalent, apparently, but played a larger role in instructional decision making. These kinds of tests apparently had considerably more validity for the teachers in terms of their suitability for students and their curriculum coverage, two prime criteria for teachers.

What other factors contributed to the teachers' use of tests? Grade level, consistent with other studies, was an important factor (see Goslin, 1965; Yeh, 1978). Less testing went on in the primary grades in the sample. More interesting, however, was the finding that the availability of classroom aides was associated with greater use of tests. It was hypothesized that aides provided a support function for the teachers--both in record keeping and in making possible instructional alternatives--that enabled teachers to use test results for decision making and to implement those decisions. During the 1978 study we saw the potential importance of making sufficient resources available to teachers to implement any new idea, and that the systematic use of test data to improve instruction, in 1978, was a relatively new idea.

Adequate knowledge and training in the use of tests, it appeared in 1978, were necessary resources. The survey indicated that most training related to testing occurred during preservice education. Thus, while younger teachers might have been exposed to newer approaches to testing, many older teachers perhaps had not. Given, in addition, the questions the survey raised about the efficacy of teachers' training in testing, the need for additional staff development activities seemed quite clear on the basis of our 1978 study.

The data from this study had an early bearing on the direction that our present investigation of test use would take.

THE EXPLORATORY FIELDWORK

Intentions

The field work was inteded, in conjunction with other Phase I planing activities, to serve two purposes:

(1) To help refine and focus the conceptual framework and research questions guiding the three-year study.

(2) To inform construction of survey instruments to be used for collecting data from the national sample of teachers and principals.

The on-site field work was designed to address such questions as: the range of ways teachers and others in schools seem to have for assessing student's performance and progress; the range of purposes that assessment results--test scores and other information--seem to serve; the kinds of assessment and uses of results that seem most pervasive, most influential for curriculum and instruction; the factors seeming to impact on assessment practices and uses most significantly; the relationships among those factors; and the adequacy of the study's conceptual framework. The fieldwork was aimed to provide information in response to such questions as these and so assist in refining and focusing the survey design.

The fieldwork was simultaneously intended to inform the construction of instruments for the national survey. The exploratory effort undertook to discover whether educators in the schools visited would find important study issues too complex, too ineffable, or otherwise too difficult to address succinctly, simply, and at the same time (from their point of view) accurately. Attention was also given to the kinds of questioning strategies and forms that seemed to bring the clearest, most precise responses to particular issues. The fieldwork also sought to identify what types (or role categories) of practitioners in schools were likely to be best informed on certain factual matters, e.g., have complete information on the school-wide testing program, know who requires that particular tests be given, etc. More fundamentally, the fieldwork aimed to comprehend as fully as possible the ways teachers and others think about and talk about the evaluation of student achievement, their instructional decisions and practices, and other matters into which the survey would inquire. In so doing, the exploratory work strived to provide data so that the language and concepts of the survey could be aligned with language and concepts through which teachers and principals organize their experience; that is, one of our prime concerns at this stage in the project reflected the issue of validity previously described, in which integration of the conceptual schemes of researcher and participant is critical.

Following from the purposes and objectives outlined above, the field work was oriented to explore issues related to the following questions:

1. What Kinds of Tests and Other Assessment Techniques are Administered?

2. What Purposes are Particular Kinds of Tests and Other Techniques Intended to Serve by Those Who Require Them?

3. What are the Features of the Social Contexts In Which Various Kinds of Tests and Other Techniques Occur?

   (including staff members' attitudes, perspectives, and reasoning on student assessment, their levels of experience and training, demographic characteristics of the school enrollment, etc.).

4. What Are the Features of the Organizational Contexts in Which Various Kinds of Techniques Occur?

   (including leadership actions, in-service programs, organization of instruction, etc.).

5. How and By Whom are the Results of Various Kinds of Tests and Other Techniques Actually Used?

Sites were selected for field work in terms of the following criteria:

1. Diversity in Required Testing Program
2. Geographic/Regional and Demographic Diversity (including diversity of ethnicity, socioeconomic status, and first language among the students served)
3. Variation in District Size and Resources
4. Variation in Local Instructional Programs
5. Variation in Reputed Skill and "Sophistication" in Test Use
6. Accessibility within Budget Limitations

Phone contacts to gather appropriate selection information were made with persons familiar with state testing programs, with the salience of testing in different regions of the United States, and with local district activities. A set of "interesting" districts was thus identified. Then, using a standard telephone protocol, information was gathered from officials in these "nominated" districts on district and school activities. On the basis of these calls, three districts were chosen.

Three schools for site visits were identified in each district with the assistance of district personnel. During this process, an effort was made to locate a roughly balanced number of elementary and high schools, schools serving higher and lower socioeconomic populations, schools with more traditional programs, and schools with more innovative instructional programs.

Exploratory field data were gathered primarily by interview. A detailed description of the interview forms and procedures followed appears in the Test Use Project Annual Report to the NIE, 1980. In brief, two forms of an interview schedule were used. We were concerned, first, with the need to balance the conceptual schemes of researcher and participant. Second, we were equally concerned with minimizing biases that might stem from the questions asked by the researcher or from the kinds of answers offered by respondents. Therefore, one form of the interview was deliberately direct and addressed matters of "testing." The second form of the measure worked by the method of indirection, and addressed matters of "information teachers use for classroom decisions." Interviews averaged 45 minutes in length. They were conducted in three school districts (one in the Northeast, one in the Midwest, and one in the Southwest) and nine schools with respondents in the following roles:

Principals	7
Vice Principals	3
Department Chairs	8
Counselors	6
Classroom Teachers	44
Specialists	7
District Administrators	4
Member of Intermediate Education Agency	1
	80 Total

The results of the field work are summarized here in two forms. First, findings across the districts and schools are presented. These findings primarily address study questions 1, 2, and 5 (see pages 29-30), which were concerned with tests administered, intended purposes, and test users. Second, descriptive narratives of each district and school are provided. These narratives, which primarily relate to study questions 3, 4, and 5, which were concerned with social and organizational contexts of testing and test users, are intended to provide an interpretive and contextual background against which to view the findings reflecting test administration and purpose.

General Findings

Across the nine schools in the three districts visited, a wide range of assessment techniques was evident. It is important to note, at the outset, that respondents referenced these almost always by their proper names or by vernacular variants of proper names. That is, they rarely talked about "norm-referenced tests," "criterion-referenced tests," "objectives based tests," "curriculum-embedded tests," etc. Instead, they spoke about "the Ginn placement," "the CTBS," "the Key Math," "that state matrix test," the "Sucher-Allred," and so on. When respondents did refer to kinds of tests, most often they gave them functional class names, e.g., "diagnostic tests," "placement tests," "pre-tests," "unit tests," "semester finals," "the competency tests." Exceptions were "standardized tests," minimum competency tests," and "district tests" (or, the "district testing program," which referred to district-developed, continuum-of-objectives-based measures in the particular sites visited).

These observations are important in that they had obvious implications for our survey instrument development. But they are also noted here to call attention to the fact that the typology of tests and other techniques used in this report is one developed by the researchers using categories salient to the practitioners interviewed.

As expected, a wide range of assessment techniques was reported by the teachers from the nine schools. These 44 teachers (22 elementary and 22 secondary) collectively mentioned the use of eight categories of assessment devices for a total of 351 citations, which is more than likely a low approximation of the actual amount. The assessment categories as well as the number of citations of assessments in that category (in parentheses) follow: Standardized tests (43), Curriculum-embedded tests (63), District objective-based tests (19), Minimum competency tests (12), School-departmental, and/or grade-level tests (17), Teacher-constructed tests (101), Diagnostic instruments (11), and Other evaluation techniques (75). The "other" category included such techniques as homework, worksheets, conferences, book reports, discussions, observations, etc.

As can be seen from the above frequencies, teacher-constructed tests and "other" evaluation techniques were cited most often by the teachers interviewed, a finding which is fairly consonant with Yeh's (1978) conclusion that curriculum-embedded tests and teacher-made tests are used to a much greater degree than standardized tests, but despite high frequency of testing, teachers are more likely to use personal observations and interactions with students than test results to assess student's progress. This latter point was not reflected in the frequencies given above but it is possible that many of the teachers, and especially those at the elementary level, failed to mention many of the informal assessment activities that occur because they are used so frequently and are so much an integral part of the teaching process. This possibility influenced the manner in which we conceived and phrased items on the survey instrument so that the subject of informal assessment could be explored further.

The amount of time these assessment techniques take to prepare, administer, and/or grade was also explored. Again, as expected, a wide range of time spent on evaluation in the classroom was reported by the elementary and secondary teachers interviewed. However, on pursuing this matter it became apparent that teachers experienced difficulty in providing an exact estimate of time indices. This was due to a variety of reasons. For one, some teachers could simply not remember how long the tests took. More commonly, it was discovered that teachers allowed different students varying lengths of time to finish the tests and thus found it difficult to average the time amounts for all students. When asked about the informal techniques they used, teachers found it next to impossible to estimate the time they spent as many of the techniques were ongoing and/or overlapping.

Although the aforementioned difficulties were encountered during the interviewing process the teachers' reports gave some indication of the time devoted to evaluation. The teachers tended to be conservative in their estimates and when ranges of time were given for a particular assessment technique, we selected the midpoint of this time frame for analysis purposes.

The analysis of the data showed that the 22 elementary teachers interviewed spent an average of approximately 11 percent of their reading and math instructional/class time assessing their students. The 22 secondary teachers reported that about 24 percent of their English and math class time was spent on evaluation. The proportion of total classroom time given over to assessment was quite large for both the elementary and secondary teachers; one to 64 percent for elementary and six to 75 percent for secondary.

At first glance it appeared on the average that the secondary teachers spent more time assessing their students than the elementary teachers. However, when looking at the responses concerning the types of assessments given, the vast majority of the secondary teachers' responses were for formal pencil-and-paper tests. Perhaps more formal testing is occurring at the secondary level than at the elementary grades because of the ages of the students involved and because the secondary teacher has less time for the use of informal techniques and/or observations. As the elementary teacher ususally spends the full school day with the same group of students, he/she has more opportunities for informal evaluations and less need for the more formal ones. Also, because the informal techniques were not cited by the teachers as frequently as the more formal ones, the difference in the percentages of time allotted to evaluation by the two sets of teachers was quite large.

The analysis also showed similar results for the total amount of time the teachers spent on evaluation. This total time includes the preparation, administration, and grading of tests/assessments. The elementary teachers reported on the average that 15 percent of their time (which includes instructional and non-instructional/preparation time) was spent on assessment while the secondary teachers spent 34 percent of their time on the same. The ranges reported by the elementary and secondary teachers were three to 56 percent and nine to 69 percent, respectively. Again, teachers' tendency not to report informal assessments and the use of many more formal evaluation techniques at the secondary level may account for some of the difference in the amount of time spent on assessment in elementary and secondary classrooms.

Range of Tests Administered

Fieldwork indicated that a wide range of tests were being administered. For example, standardized tests, such as the Comprehensive Tests of Basic Skills (CTBS), the Metropolitan Achievement Test (MAT), Iowa Test of Basic Skills and of Educational Development (ITBS, ITED), etc., were administered in each school district visited.

Curriculum-embedded tests of various types were also given everywhere, but almost exclusively at the elementary grade levels. Most of the curriculum-embedded tests accompanied commercially-produced, elementary-grade series in math and reading. Among those given frequently were placement tests; the "unit" or "criterion" tests designed to assess achievement on a specific portion of the curriculum; and "end of the book" tests (i.e., those the student took at the completion of a given reading or math "level").

Minimum competency tests were given in two of the districts. In one case they were district-developed and included four separate instruments assessing fundamental math skills and four assessing skills in the language arts. These tests were given at the high school level and passage of all eight was required for graduation. In the second district, an instrument developed by the state for administration to ninth grade students included the general domains of reading, mathematics, and writing. Its function was only diagnostic.

A statewide assessment measure was given annually in one district to a matrix sampling of students at certain elementary and high school levels. Individual student scores were not reported to schools, but aggregations by grade-level, school, and district were provided on various subskills in reading, mathematics, and writing.

District tests, district-constructed and mandated for use district wide, were part of the assessment picture in two of the three districts visited.

School-, departmental-, and/or grade-level tests were found in five school sites. One high school, for instance, had just developed and administered a writing sample in all grade levels. Departments in several high schools had teacher-developed mid-terms and finals for particular courses. And in two elementary schools in one of the districts, teams of teachers at particular grade levels constructed and gave common tests keyed to their social studies curriculum.

Diagnostic instruments were also employed largely, but by specialists such as remedial reading instructors, teachers of the "learning disabled" and "emotionally handicapped," and Title I program staff members. Almost all of these were found in elementary schools.

Teacher-constructed tests, quizzes, and the like were, of course, extant in every site.

Other measures of student achievement were also prevalent in all classrooms. In the elementary grades, students' daily worksheets, classroom performance, along with homework and other assignments, were mentioned as ways of evaluating students' progress. These same types of "measures" were among those used by high school teachers. The latter also cited conferences with students, peer evaluation of classroom reports, oral quizzes and question-answer sessions, group discussions, and a wide variety of written assignments as assessment techniques.

The specific configuration of tests being administered in each of the districts visited is provided in the district narratives.

Range of Reported Uses

Distinct patterns of use also grew out of fieldwork analysis, which suggested that test scores and other assessment results were used for a finite number of purposes across the sites visited. At the classroom level, there was little school-to-school or district-to-district variation in the range of uses respondents' reported. Eleven types of uses for assessment information were inductively derivable from the specific comments of educators interviewed. Recall that the uses listed below are those which individual respondents said they themselves made of test scores and other student assessment "data."

(1) Referral to and/or placement in special programs, appropriate classes, appropriate "tracks," etc.

(2) Within-classroom placement of students at appropriate levels in individualized programs, in reading or math groups, in occasional, temporary skills remediation groups, etc.

(3) Planning instruction: "figuring out my class' strengths," "learning what the group needs," "getting feedback so I know what we have to go over again," "working with one of my grade-level groups of teachers to decide what areas they need to strengthen," etc.

(4) Monitoring student's progress, "seeing how they're doing as we go along," "just getting a sense of whether they're learning anything."

(5) Holding students accountable for doing assigned work, maintaining class discipline.

(6) Assigning report card grades.

(7) Certifying students' competency for promotion, high school graduation.

(8) Counseling and advising students about how they are doing, about their preparation for future courses and academic goals, about their achievement, motivation potential, etc.

(9) Informing parents of how their children are doing in regularly scheduled conferences, at "back-to-school" nights, special meetings, when problems arise.

(10) Reporting to higher organizational levels within the district --to the principal, district office, the school board--on student achievement.

(11) Comparing groups of students with others, judging how a class, school or district is performing relative to others.

Patterns of Assessment Results Use

From the respondents' comments about how they used the results of particular tests and other assessments we developed a coding scheme to index the importance of particular results for particular purposes. This simple scheme depicted the use of a score or result for a given purpose as: (1) the sole information source used; (2) one of two or three major sources; (3) one of many sources; (4) a verification source, i.e., used ancillarily to check decisions or conclusions already reached based on other information sources, and (5) not used, simply administered.

Interview data from the 44 classroom teachers included 330 descriptions of how the results of particular types of assessment were used.* They also included 21 statements that the respondents did not use results of types of measures that they administered.

---

*Redundant uses for different tests of the same type were dropped out in collapsing the 346 tests/assessment means cited into the eighttypes of assessment listed earlier in this section.

As Table 1 indicates, teachers rarely used only one type of assessment information to make a given decision or accomplish a given purpose. Only 5.1 percent of the uses cited (including statements of non-use) were "sole source" uses, i.e., results used alone to make a given decision. In two-thirds of the cases, results from a particular type of assessment were used as one among many types of information employed for the particular purpose at hand.

Table 1

Overall Patterns of Assessment Results Use

	Functional Importance					
	Sole Source	One of Several major Sources	One of many Sources	Verification Source	Not Used	Total
Instances Mentioned	18 (5.1%)	65 (18.5%)	237 (67.5%)	10 (2.8%)	21 (6.0%)	351 (100%)

In short, it appeared that teachers were most likely to look at a variety of different kinds of information as they make the judgments, analyses, and reports they must make as part of their routine professional activities.

Test information used as sole and major criteria: If most means of assessment provide information that is used jointly with others, which means do seem to provide information that functions as a sole or major criterion in teachers' activities? Table 2 provides an answer in overview.

Table 2

Types of Tests Used by Teachers as Sole and Major Sources of Information for any Purposes

Test Type	Total* Citations all levels	Count (Column %)		Total: Sole & Major (% total in Table)
		Sole Source	Major Source	
Standardized	43	6 (33.3)	5 (7.7)	11 (13.2)
Curriculum Embedded	63	5 (27.8)	12 (18.5)	17 (20.5)
District Objective-Based	19	1 (5.6)	6 (9.2)	7 (8.5)
Minimum[+] Competency	12	0 (0.0)	0 (0.0)	0 (0.0)
Statewide Assessment	10	0 (0.0)	0 (0.0)	0 (0.0)
School/Department Grade-Level	17	0 (0.0)	9 (13.8)	9 (10.8)
Individual Teacher-Constructed	101	5 (27.5)	15 (23.1)	20 (24.1)
Diagnostic	11	0 (0.0)	0 (0.0)	0 (0.0)
Other	75	1 (5.6)	18 (27.7)	19 (22.9)
TOTALS	351	18	65	83 (100.0)

*Count of all instances in which test type was mentioned as used in any way, including "not used" category.

[+] Minimum competency tests were used as the sole source for deciding whether students graduated from high school in one district, but this decision was not made by classroom teachers or other school-level practitioners.

From the above, a picture began to emerge of teachers drawing upon many types of assessment to do their routine instruction-related work. And the fieldwork data suggested that the types of assessment they use most frequently in this routine work tended to be those that are

- most immediately accessible to teachers and which provide most immediate results; those over which they have most control--can administer when they choose and can see the results promptly;
- those which purport to serve functions isomorphic with the tasks teachers must routinely do; i.e., curriculum-embedded placement tests figure significantly in placement decisions; records of progress through a continuum for placement in a continuum; tests that teachers design or text publishers produce for measuring achievement on a unit of instruction for monitoring progress and grading students on that unit, etc.
- those which teachers deem to "cover" most exactly the content of the material they are teaching.

In short, those tests teachers see as linked most closely to the routine, practical activities of their everyday professional lives are those they use most often. Additionally, the phenomenological evidence of everyday experience with students plays an important role in teachers' assessments of them.

The single exception to this generalization appears to occur in the use of standardized tests. For the most part, teachers used these for general reference, to get an initial sense of how their new classes "look" relative to others, or as a normative reference point against which to gauge progress--except, it seems, when they are required to do otherwise by district mandate.

Test information that is not used: In 21 instances, teachers said they did not use the results of one or another type of test that they gave. Ten teachers mentioned their non-use of standardized test results;

seven mentioned non-use of statewide assessment. In the case of the latter, teachers had no access to students' individual scores or results aggregated by class.

The above descriptions began to indicate some of the activities in which assessment results play a definitive or major role. Table 3 provides a comprehensive picture of the purposes for which they do so.

Table 3

Purposes for Which Teachers Use Various Types of Assessment Results as Sole and Major Information Sources

Count: Number of Citations

Purposes	Sole	Major	Total	(% Table Total)
Planning Instruction	1	9	10	(12.1%)
Referral/Placement: Special Program	4	5	9	(10.8%)
Within-Class Grouping and Individual Placement	7	18	25	(30.1%)
Holding Students Accountable for Work, Discipline	1	6	7	(8.9%)
Assigning Grades	0	9	9	(10.8%)
Monitoring Students' Progress	0	6	6	(7.2%)
Counseling and Guiding Students	5	8	13	(15.6%)
Informing Parents	0	1	1	(1.2%)
Reporting to District Officials, School Board, etc.	0	2	2	(2.4%)
Comparing Groups of Students, Schools, etc.	0	1	1	(1.2%)
*Certifying Minimum Competency	0	0	0	(0.0%)
TOTAL	18	65	83	

*Note: In one district visited, tests of minimum competency were required for high school graduation. Respondents, however, took this as obvious and rarely mentioned that they served in this way. When they did speak of the uses of minimum competency results, they described their uses for other purposes.

As Table 3 shows, test scores seemed to play an important role in student placement decisions. In 40.9 percent of the instances in which teachers reported that they used assessment results as a sole criterion or a major criterion, the placement of learners was at issue. The use of scores as a major basis for in-class placement was especially frequent.

Summary. Most often, teachers seemed to consider the results of several types of assessment collectively in arriving at a particular decision or carrying out a particular activity. When they reported departing from this practice, it was more often in the direction of weighing test scores more heavily than in the direction of counting them less. (Citations of results as sole and major information sources equaled 23.6 percent of the total; citations of results not being used or used only in verification equaled 8.8 percent of the total.) The placement of students seemed to be an activity in which the results of one test or type of test may count more heavily than in others.

Relationships Between Types of Tests and Categories of Use

Table 4 summarizes the test type/use type relationships reported by both the elementary (n=22) and secondary (n=22) classroom teachers interviewed. The table indicates that the main uses of test and other assessment results include:

- Planning for instruction
- Grouping students and placing them at levels of individualized programs within classrooms
- Grading
- Monitoring students' progress, i.e., keeping track of how they are doing over time.

Table 4

Types of Tests and the Uses of Their Results

USES Counts: Elementary Secondary Cell Total	Standardized	Curriculum-Embedded	District Objectives-Based	Minimum Competency	Statewide Assessment	School, Department, Grade-Level	Teacher-Constructed	Diagnostic (Skills)	Other (Informal)	TOTAL
	Type of Test									
Planning Instruction	9 4 13	8 2 10	3 0 3	1 3 4	2 0 2	1 2 3	11 13 24	1 1 2	13 8 21	49 33 82
Referral/Placement:	9 2 11	0	0	0 1 1	0	0 2 2	2 1 3	0	2 4 6	13 10 23
Within Classroom Grouping & Individual Placement	4 0 4	18 0 18	5 0 5	1 2 3	0 1 1	1 3 4	2 4 6	6 0 6	11 3 14	48 13 61
Holding Students Accountable for Work, Discipline	0	3 0 3	0	0	0	0	4 4 8	0	2 0 2	9 4 13
Assigning Grades	0 1 1	14 3 17	1 0 1	0 1 1	0	0 5 5	15 17 32	1 0 1	7 1 8	38 28 66
Monitoring Students' Progress	0	14 0 14	4 0 4	0	0	0 2 2	10 8 18	1 0 12	10 2	39 12 51
Counseling & Guiding Students	1 2 3	0	2 0 2	0	0	0	2 8 10	1 0 1	4 2 6	10 12 22
Informing Parents	0	0	1 0 1	0	0	0	0	0	1 0 1	2 0 2
Reporting to District Officials, School Board, etc.	0	1 0 1	2 0 2	0	0	0	0	0	3 0 3	6 0 6
Comparing Groups of Students, Schools, etc.	1 0 1	0	1 0 1	0	0	0	0	0	1 0 1	3 0 3
Certifying Minimum Competence	0	0	0	0 1 1	0	0	0	0	0	0 1 1
TOTAL Use CITATIONS	24 9 33	58 5 63	10 0 19	2 8 10	2 1 3	2 14 16	22 55 101	10 1 11	51 20 74	217 113 330
Explicit Statements: "NOT USED"	5 5 10	0	0	1 1 2	0 7 7	1 0 1	0	0	0 1 1	7 14 21
Total Citations	29 14 43	58 5 63	19 0 19	3 9 12	2 8 10	3 14 17	46 55 101	10 1 11	54 21 75	224 127 351

Summary. The exploratory fieldwork indicated that the sample teachers most frequently drew on the results of three types of assessment. These are (1) their self-constructed tests, quizzes, and written assignments, (2) other assessment techniques that they devised or chose to seek out and use, such as class discussions, peer evaluations of work, conferences with students, talks with their students' previous teachers, oral reading sessions, etc.; and (3) curriculum-embedded tests--those that come with district-made curriculum "packages" or commercially publised texts, kits, and the like. They appeared to use each of these three types especially, but others as well, in accomplishing a variety of purposes. That is, teachers seemed to refer to each kind of assessment result for making a variety of judgments, just as they seemed to make a given decision by referring to a variety of assessment results. Principals seemed to engage in a similar practice, although the test scores they used most often and the purposes for which they used them most frequently differed from those of teachers. All this suggested, of course, that the national survey should examine patterns of test type/test use relationships. It should not assume simple one-to-one correspondences between a test score and a use.

Teachers most frequently cited test scores and other assessment results as serving them in four activities: planning instruction, grouping and placing students in a continuum of objectives within the classroom, assigning grades, and monitoring students' progress over time. Counseling, guiding, and other use seemed to follow from the factors previously discussed.

A final point is worth noting again. Returning to Table 4, it is obvious that some activities for which teachers use student assessment results are relatively "under-mentioned." For instance, conferences with parents are a routine part of teachers' work, especially at the elementary school level. A talk with any teacher about his/her students inevitably includes comparisons with students in other classes or schools, students in previous years, and so forth. That these activities were cited relatively infrequently as uses of assessment results was troublesome to us. In talking with teachers, however, it became evident that many of the practical tasks for which teachers use test information are, in fact, "transparent" to them. That is, they are so much a part of everyday life that they go un-noticed. They are treated, literally, as unremarkable. That this is so is probably best illustrated by a comment made by a high school assistant principal in the first district visited, who explained in the same breath that they did not pay much attention to CTBS scores in his school because the typical freshman entering the school was "two years, at least, below grade level."

This should serve as a caveat that Table 4, and the discussion which has followed from it, is not a complete picture of the frequency with which the teachers interviewed use test results for certain purposes. But, given the open-ended nature of the interviews, it is very likely a comprehensive picture, overall, of the kinds of uses that test and other assessment results serve.

District Narratives

Content analysis of the taped transcriptions of the nine schools across the three districts provided information bearing on the social and organizational contexts in which tests are administered and used. This analysis suggested that five factors seem to have a bearing on the atmosphere in which tests are administered, and consequently how they are "valued" by teachers and used/not used in classroom decisions. On the basis of fieldwork, these factors emerged as:

(1) state testing policy and requirements

(2) coherence of school/district testing policy and requirements

(3) leadership in the instructional uses of assessment information

(4) locus of ownership of the assessment program

(5) recognition that no single test can serve (nor is intended to serve) the information needs of decision makers who reflect a variety of interests from broad program accountability to specific classroom practice.

While we had not intended fieldwork to provide a picture of "exemplary" test use (that would possibly emerge during Phase II of the project), analysis of responses did suggest a tentative picture of how contextual factors may converge to make tests appear "usable" (as previously described on pages 6, 42 of this report). As will be seen later, the district which seems to have a successful testing program--successful from the standpoint of reconciling or balancing external testing requirements with school-level uses of testing--assumes an organizational posture which has elements of centralism and diffusiveness. The importance of this observation emerged from our cross-project collaboration. That is, one of the Test Use Project staff was involved with CSE's Evaluation Design Project, which has been examining evaluation/testing matters at the district level. Part of the collaboration involved the production of a CSE monograph entitled Evaluation in School Districts: Organizational Perspectives (Bank & Williams: 1981, in press). During this inter-project work, some of the findings stemming from work at the district level, and which are discussed in the monograph, took on importance for an investigation of testing at the school and classroom level.

For example, it is possible that an organization and its constituent parts can (or perhaps, should) be "loosely-coupled" in some regards and more tightly coupled in others. This variable posture, when applied to our fieldwork findings, appears to lend itself to multiple uses of assessment information: uses which are central and concerned with external accountability and reporting requirements and uses which are spread out and reflect the decision needs of individual schools and classrooms. This is not to suggest that a balance of central authority and dispersed decision making is the only approach that will lead to development of a"usable" testing program. But it appears to be the approach that has evolved, over time, in one of the districts we studied, and it seems to reflect not only organizational reality but the careful determination of various decision needs and specification of an assessment information system that will meet these needs.

Assessment programs often intend to provide information for use at local, state, and/or federal policy levels. Often the program will tend to emphasize the information needs of one of these levels to the exclusion of the others. Many assessment programs appear to be driven, or are perceived by the people in them, to be driven more by broad, external accountability than by concerns for classroom and school-specific information. (This issue of external "linkages" is also discussed in Bank & Williams, 1981). Audiences associated with these external requirements often ask for assessment information that can be used to compare educational programs rather than to show the growth of individual pupils in terms of a specific set of educational objectives. A school system which tends to respond more to the external audience than to others frequently relies on the collection and analysis of pupils' scores on a norm-referenced test. It may be criticized for lack of concern with individual students and their growth on precise instructional objectives. A school system tending to respond solely to audiences concerned with individual student growth in a given classroom (no such system was discovered in the present study) might tend to rely more on criterion-referenced or objectives-based tests to provide information for diagnostic and prescriptive information. A school system taking this position might be subject to questions about the educational significance of the scores obtained on this kind of test -- What do they mean? Do they show whether the learning that has taken place is important or trivial? How do the scores obtained on these tests compare with the scores obtained on other kinds of tests?

A school system might attempt to reconcile both kinds of information needs, to examine the operant assessment requirements, to investigate their own assessment needs, to determine which kinds of information will address the range of needs, to decide which kind of measure is most appropriate for generating the information addressing a particular decision area, to specify for its participants the intended uses of various measures, and thus design a coherent assessment program which is perceived to have a variety of overlapping uses.

One of the districts we spent time in appears to have developed this kind of assessment program. The two other districts we visited seemed to be trying to move in this direction, but still seemed to be more concerned, or at least their teachers felt they are more concerned, with external accountability issues.

District One

This school district, located in the urban Northeast, has 24 elementary schools (kindergarten to grade 6 primarily; a few are K-8), 2 middle schools (grades 7-8), and 3 high schools (grades 9-12). Total enrollment is 27,000, with approximately 50% Black, 30% Hispanic, and 20% Anglo and other combined. The district had approximately 18 schools that are Title I eligible.

The state in which this district is located has a minimum competency testing program which is still in a formative stage of implementation. While no final determination had been made at the time data were collected, school district officials did not anticipate that the proficiency test would become a requirement for high school graduation. By the provisions of the state requirement, which focuses on "education, evaluation, and remedial assistance," all 9th graders are tested for proficiency. Any student scoring below a certain cut-score (established by the state) must receive remedial assistance from the local school/district. The state required testing covers the areas of reading/language arts, mathematics, and also calls for a student writing sample.

Beyond the state required minimum competency testing program, the district has its own testing program, which is also in a formative stage of development. This district testing program deals with the areas of reading and communication arts, and includes the use of a locally developed criterion-referenced measure. This test is structured by grade, scope, and sequence, is intended to provide mastery data, and is administered by teachers and/or reading consultants. It becomes part of the student's permanent school record and follows him/her from grade to grade and school to school. District officials anticipate that when this test has been fully developed, it will become part of the district's response to the state required minimum competency testing program.

As part of the district's required testing, the Metropolitan Achievement Test (MAT) is used in grades 2 through 8. It is administered every spring. At the high school level, the Comprehensive Tests of Basic Skills (CTBS) is administered in the 11th grade.

The district test, which is accompanied by a specific curriculum, is supposed to be administered in all schools as part of an attempt to standardize the curriculum; this was apparently not happening in actual practice, however.

District Two

The second district we visited is located in an urban area in the Southwest. This district has over 100 elementary schools, 20 junior high schools, and 14 high schools. Total district enrollment is a little over 100,000.

The state in which this district is located has a required minimum competency program for high school graduation. Local districts can use a state developed test or select/develop their own. This district has developed its own competency program to meet the state requirement. Among the tests in use in elementary schools are: CTBS; the state assessment program; the district competency test; and variable use of a range of curriculum-embedded tests and teacher observation and classroom interaction. Among the tests in use in the high schools are: the state assessment program; district competency tests; CTBS: tests associated with college entrance; and variable use of teacher constructed measures and classroom observation and interaction.

District Three

The third district visited, which demonstrated multiple and "exemplary" uses of assessment information, is located in a rural community in the Mid-west. This district has seven elementary schools, three junior high schools, and one high school. Total district enrollment is a little over 5,000 students, of whom only 6 percent are minorities.

The state in which this district is located has no required minimal competency or proficiency testing. The only state requirement is that districts must identify student needs and set plans to meet desired levels of achievement.

Among the tests used are the Iowa Tests of Basic Skills (ITBS, grades 3-8), the Iowa Tests of Educational Development (ITED, grades 9-12), the Cognitive Abilities Tests (CAT, grades 1,3,6, and 9), district/school developed objectives-based tests, and curriculum-embedded tests.

Schools in this district also enjoy the resources of an Area

Education Agency (AEA). One of the functions of this agency is to provide technical assistance to schools and individual teachers who have questions, problems, and needs in testing.

This district differs from the first and second on some important dimensions. In the third district, the fairly well accepted, district/school developed tests seemed to reduce the amount of time that teachers spend constructing and administering their own tests (especially at the elementary schools), thus freeing instructional staff for other tasks. These locally developed tests are largely seen as complementing the use of standardized tests, and serving different, though related decision needs. In addition, with greater acceptance of district testing there seemed to be a clearer sense among the teachers of both the "district" itself as an educational system and its testing policy and intentions, which teachers did not seem to see as threatening.

Much of the information provided by the respondents seemed to reflect needs, issues, and concerns about the levels of decisions (Baker, 1978) that might need to be made on the basis of assessment information. Two of these, levels 1 and 2, were alluded to previously. Level 1, reflecting information needs to make decisions about individual students, is of prime concern among teachers, specialists, guidance counselors. Level 2, reflecting information needs to make decisions about groups of students within a school, is also of concern for teachers, but somewhat more so among department chairpeople, grade level coordinators, and principals. Level 3, reflecting information needs to make decisions about groups across schools, is the concern of decision makers at LEA, SEA, federal levels, and the general public.

Test Uses/Issues in District One

In one of the schools in this district, an elementary school, respondents did not appear to value the district testing program. There was an impression that the school administration, which had been recently appointed, was selected to stress the district program and the need for accountability at the level of the school. Respondents seemed not to see the purpose nor the relevance of this testing program. They did seem to be concerned with the kinds of tests available, their match with classroom curricular concerns, and the instructional unit at which the test has decision-making relevance. Teachers here were largely concerned that the tests being used did not seem to match their instructional concerns and related information needs. They saw little coherence in the district/school testing policy and expressed little confidence in its classroom use.

In another elementary school in this district, the school administration and some of the curriculum and resource specialists seemed to concern themselves to an extent with accountability (level 3) decisions, but the teachers did not seem overly concerned with this state of affairs. It appeared that they not only went about the business of making their in-class and in-school (level 1 and 2) decisions, but also received a level of expert assistance in making these decisions that was not encountered in the first school.

The third school visited in this district was a high school. Perhaps the most severe problem at the school is the fact that most of its students do not graduate. In an attempt to specifically pinpoint student deficiencies and make appropriate curriculum changes, the norm-referenced test being administered -- the CTBS -- had not proved useful. There was a hope among staff that the district testing program (as well as improved use of department tests) would come to serve as student motivators and as a means to restructure the curriculum.

District Summary

Several testing issues emerged in this district. First, the state-required testing program was still in a formative stage. The district testing program, which responded to state competency testing, was equally recent. The district program seemed intended not only to serve the needs for competency testing but also to help standardize the curriculum district wide. At one school it was seen by teachers as no more than another accountability measure; if it had some instructional value, it was not seen by the teachers. In this school, teachers seemed to have little sense of district, or school, testing policy. Teachers seemed to feel that required testing served only level 3 decisions; it helped them not at all with level 1 and level 2 decisions and, indeed, may get in the way of teachers using measures of their own choice for these purposes.

In the second school, teachers seldom mentioned the district testing program. The teachers here perhaps understood the purposes of the program and so felt less threatened by it. On the other hand, they simply may not care either way if it does not get in the way of their classroom activities. One explanation is that concerns of the district testing program (and level 3 decisions) are seen in this school as the responsibility of the school administration and specialists. It appeared that these specialists, some of whom were concerned about the amount of testing taking place, used the district measure not only for district concerns but also, where appropriate, to help classroom teachers with their internal level 1 and level 2 decision.

In the third school, standardized tests administered in the past had served no purposes in instructional improvement. There was a distinct impression that the school was assuming a policy of "wait and see" in the hope that the new testing program would help them.

In general, the district testing program seemed to suffer from lack of clear policy and guidelines; in only one of the elementary schools was there any sense of leadership in the instructional use of assessment information. It seemed that at the high school a policy was emerging which may lead to a sense of ownership of the testing program.

Test Uses/Issues in District Two

In one of the elementary schools in this district, a prime concern of the teachers was that tests would be used not only to monitor building progress, but also to evaluate teacher performance. The principal stated that if teachers believe they will be evaluated on the basis of test scores, this is acceptable if that is what is required to achieve instructional improvement.

In the second school visited, a high school, the impact of minimal competency testing and the time devoted to this testing has had a profound influence both on teacher attitude toward required testing and also toward the uses they make of other kinds of tests.

In the third school visited, also a high school, the impact of minimal competency testing was felt to be equally high, influencing not only the amount of testing taking place but also the content of instruction in the classroom.

District Summary

The advent of minimum competency testing has had an observable and, from the standpoint of some respondents, a negative effect on regular classroom instruction and the kinds of resource options made available to teachers. While the effect seemed to be more pronounced at the high schools, it also had a bearing on the policies of elementary schools visited.

In many respects, teacher concern for amount of testing, kinds of tests administered, and the uses to which they are put echoed the kinds of responses encountered in the first district visited. This is especially true with respect to minimal competency testing.

Test Uses/Issues in District Three

In one of this district's elementary schools, while there were some teacher-perceived problems with testing, teachers seemed to view tests as a more useful decision-making tool than was the case in the first two districts. The test selection/development/use inservice offered in this district appeared to strongly influence teacher acceptance and use of test results. Of equal importance, however, are the services offered by the AEA, a kind of teacher center in which advice, technical assistance, and actual tests can be constructed/selected by teachers.

Another factor appearing to influence teacher use of tests was the atmosphere in which testing policy is conveyed. The district and school administration apparently set broad test information requirements, intended to serve both external accountability and internal instructional improvement needs, in which departments and teachers have several options.

One of the respondents in the first school visited described the history of the district's approach to testing and the role of centralized training and technical assistance. As a media specialist responsible for providing "teachers with the materials they need to teach kids," several years ago he developed an interest in computer assisted instruction. His interest in CAI led to using local computer services for test scoring and data analysis. This led to a district interest in "computer analysis rather than hand scoring, to give you a better idea (of) where the kids are ... You don't have the time or the expertise in the classroom, generally, to do that; the computer does it in one fell swoop." This quick and accurate scoring service, covering all the various kinds of tests used, was available to any teacher in the district. Over the years, further, the link from CAI to test scoring and analysis led to a further computer application. That is, teachers had gradually developed large banks of educational objectives, had written or adapted hundreds of test items written at varying levels of difficulty, and could resort to the computer files to call out a particular kind of test for a particular instructional purpose. Over the years it appears that local teacher involvement, with technical assistance and leadership from the AEA and district officials, has led to a greater degree of test sophistication and test use among teachers than was the case in district one and two schools.

Therefore, while some teachers expressed concerns about such problems as the lateness of receiving the results of the standardized test as well as their relevance for some classroom objectives, these criticisms did not carry over to testing in general. Indeed, some of the tests used were seen as invaluable for both teachers and students. Tests seemed also to be used as instructional motivators whose results were discussed by teacher and students as one more source of diagnostic information. The link between testing policy and test use seemed clearer than in the first two districts. In the third district teachers seemed to feel the testing program was in part their own, to be used for their level 1 and 2 classroom decisions as well as for school and district accountability matters, and to be tempered by teachers' professional interactions with their students.

The second school visited, also an elementary school, appeared similar to the first in terms of uses of assessment information. The norm-referenced test in use -- the ITBS -- did not appear to receive a great deal of emphasis for classroom decisions, although it was useful to the administration in making decisions about building-level effectiveness.

District developed and validated tests did appear to be weighed heavily for certain kinds of within-class decisions as well as for teacher self-monitoring. For many of these decisions, further, teachers also relied on less formal means of assessment in the interests of making the best instructional decisions.

The third school visited was a high school. Here some of the school staff interviewed seemed knowledgeable (in some cases, almost expert) in matters of testing and test use, especially in the math department. Indeed, the school administration expressed hope that the model of the math department would eventually transfer to other departments. To be effective, however, they believed that this must occur naturally with no direct interference from the administration.

In this school, the principal and associate principal emphasized the crucial role of the district in sponsoring within-school and centralized opportunities for technical assistance in testing. This school also seemed to exemplify the best uses of certain kinds of tests. In terms of the ITED, its use, as seen by the school administration, was expressed as follows: "We need at least one outside measure, something outside of our own control ... so we can just have a benchmark ... that we can compare with" in terms of school-level performance. Beyond that, item analysis of ITED scores might lead to discussion between the associate principal and a department chair if test score trends, over time, were consistently poor in certain areas. "Should this indication lead to course modification? Adding something to instruction? Do instructors want to add this area to instruction? Do they want to leave it out because they don't think it's important?" This kind of discussion indicated a measure of department autonomy or, at least, negotiated decision-making.

In this school in general, and in the math department in particular, the school-developed measures appeared to be accepted and used by teachers. Departmental autonomy in testing and the inservice and technical assistance made available appeared to have stimulated local development of tests that are quickly accessible, fit teachers' practical needs, and have high content and classroom relevance. Standardized tests were primarily used by the school administration, and seemed to be viewed neither as a threat nor as an unnecessary burden by the teachers.

District Summary

This district clearly had a different approach to testing and testing policy than the first two. It appeared that the district establishes broad policy for schools, and the schools, in turn, set broad policy for the instructional teams in the elementary schools and the departments in the high schools. Test administration, quality, and level 1 and 2 uses were also focused at the level of team or department. In addition, both the district central office and staff of the AEA provided active leadership in the development of tests and their instructional uses. Policy was clear, though flexible; it seemed to reflect an organizational system whose units could "couple" or "de-couple" as described in Bank and Williams (1981). A great deal of the testing appeared to be "owned" by the school unit of concern -- team or department. While teachers seemed less likely to rely greatly on the ITBS and the ITED, counselors were available to help interpret these scores and place them in the larger assessment context for individual teachers.

Teacher knowledge of tests and testing appeared to be greater than in the first two districts. There also appeared to be more inservice and there was certainly much more technical assistance available in the third district. This has led to the development of tests of higher quality which apparently have marked instructional relevance for the teachers. The testing situation appears to come close to the teacher's ideal (as we described it on page 16). That is, the overall testing program:

- offers tests oriented to classroom teachers
- permits teachers to use tests so as to meet their practical activities and exigencies
- does not force teachers to emphasize tests which do not fit their practical demands
- permits teachers to administer/use a variety of tests
- is sensitive to the exigencies of teaching-as-practiced.

In this district, further, the merits of different kinds of measures were not discussed by the participants in an adversarial setting. Instead, the teachers, principals, and district officials seemed to accept the need for and value in generating information that will paint the big (norm-referenced) picture, that will provide a wide-angle view about groups and programs. They did not over-emphasize this picture. They also accepted the need to generate information about the individual students and classrooms (criterion-referenced or objectives-based) that together make up the big picture. They did not over-emphasize the value of this picture either.

They seemed to be using the right kind of test to get the larger aggregate picture, and a series of other, equally appropriate measures, to get a variety of snapshots with a closer focus and with greater detail, of the separate parts of the picture. The district, the central figure, has supplied the camera -- the means to get the different pictures -- and takes the kind of shot with the degree of resolution it needs. The schools and classrooms use the same camera, but they select a kind of film that meets their needs, and then choose an angle, focus, and degree of resolution sensitive enough to get the series of shots that they need. The end result seemed to be a montage reflecting different degrees of instructional progress among different aggregates of students at varying points in time.

As with other activities stemming from our test use planning work, information collected and analyzed seemed to clarify the most critical areas to pursue in our national survey, as well as the manner in which to pursue these areas.

The next section of the report discusses the manner in which the national sample was selected and presents the results of questionnaire data collection and analyses.

THE NATIONAL SURVEY

Sampling Methodology

As mentioned previously, we intended for the survey to be national in scope, to provide both descriptive and inferential data relating to a series of practical and policy matters, and were guided by our planning work as we conceived of the design for the survey, drafted questionnaires, and considered the sampling plan. The sample had to be selected as to obtain a national picture of the uses of achievement testing, and we had limited resources to do this. Teachers were the primary target of the survey because they conduct a great deal of achievement testing and are therefore in one of the most strategic positions from which to judge the relevance of testing programs in terms of criteria we have alluded to throughout this report. In addition, to collect confirmatory data and information on relevant contextual variables, principals of selected schools and district testing officers were also selected as study respondents.

The Test Use Project's earlier fieldwork had demonstrated that frequency and uses of tests vary with grade level of students. The survey therefore included the fourth, sixth, and tenth grade levels. (Rationale for the selection of these grade levels has been provided in earlier Test Use Project reports.)

Because the focus of much testing is in the basic skills areas, the study targeted assessment in reading, language arts, and mathematics. Elementary school teachers were asked questions pertaining to both reading/ language arts and mathematics assessment. At the tenth grade level, language arts (English) and mathematics teachers were asked about assessment in their respective fields.

The survey was directed to two elementary and two secondary schools in each selected school district; with two fourth and two sixth grade teachers in each selected elementary school, and two language arts and two mathematics teachers in each selected secondary school. The target was about 400 teachers of each type. Since many districts have only one secondary school, it was necessary to sample in excess of 100 districts to meet this objective.

The sample was selected to be sufficiently representative of the target populations as to generalize to these populations throughout public schools. Factors that guided the selection included the district's minimum competency testing status, student enrollment, SES, geographic region, and metropolitan status.

Because the data collected in the study are being used to provide the basis for inferences about the influences of various contextual factors, the project was careful to design a sampling plan that would obtain general representation over the variables of interest. The conception, development, and refinement of the sampling plan proceeded as follows.

The Initial Plan

The initial conception of the plan was to draw a sample of approximately 100 districts to yield a total respondent sample of 2,100 individuals. Allowing for the inevitable shrinkage which occurs with the use of mailed questionnaires, this size of sample was considered to be an adequate basis for inferences about the nature of achievement tests in current classroom use. An illustration of kind of respondent by school district is seen in Table 5.

Table 5

Number of Respondents by Type for each District

Respondent Classification	Number for each District
District Testing Officer	1
Elementary School Principals	2
High School Principals	2
Fourth Grade Teachers	4 (2 in each of 2 schools)
Sixth Grade Teachers	4 (2 in each of 2 schools)
Tenth Grade Teachers, language arts	4 (2 in each of 2 schools)
Tenth Grade Teachers, mathematics	4 (2 in each of 2 schools)
TOTAL	21

The initial design called for a proportional probability sampling (PPS) strategy to draw a sample of districts and schools with a probability of being selected proportional to their size and representation in the population by state testing criterion (i.e., state assessment and minimum competency testing status).

During this stage of our thinking, we considered the factors that might be used to stratify the population of districts for sampling purposes; e.g., presence of minimum competency testing, size and location of district, etc. In that some strata would have no comparative usefulness in the study, our interest in them was limited to ensuring that interesting population features would be present in the sample in proportion to their representation in that population. We felt, at this time, that the most direct manner of obtaining a sample of districts was to array them in a nested ordering representing the specific characteristics of interest, and then to sample them with a probability proportional to their sizes.

The major features in this initial conception of the sampling design were the minimum competency testing matrix and geographical region. Districts were to be ordered in the cells of the MCT matrix by geographical region. The districts would not need to be ordered by size because the PPS scheme would select them in a fashion properly representing this variable.

Within districts, schools were to be selected by randomly drawing one low SES and one middle or high SES elementary school from a list of such schools and by similarly drawing the two secondary schools. At the elementary school level, lower SES schools were defined as those receiving ESEA Title I funds, and higher SES as those receiving no compensatory education funding. Aid to families with dependent children (AFDC) was used to define SES at the high school.

Because a great many of the districts in the United States are too small to have two elementary schools or (especially) two high schools, districts would be selected with probability proportional to size, so as to reduce the likelihood of drawing a very small district.

The teachers would be selected by drawing two teachers at random from a list of target teachers at each desired grade leve. At the elementary school level, target teachers were defined by grade level taught. At the secondary level, target teachers were defined as those teaching the subject area classes with greatest enrollment at their school; i.e., the most common mathematics and language arts classes for tenth graders. If time pressures could not be controlled, an algorithm for making such a selection would be described to the principal, who would then make the selection.

The samples obtained by this method would represent the responses of the "typical" teacher or principal, in that larger places (employing more of these people) would be more likely to appear in the sample than would smaller places, while the probability of including specific districts would be inversely related to their size. The net result would be that all teachers would have about an equal chance to be selected into the study; the same would be true of principals.

Respondents in districts offices, however, would not have equal probability of selection, and if their responses were to be analyzed without weighting for selection probabilities, they would represent responses characterizing the environment of the "typical student" (as if we had selected the district officers with equal probability and weighted for district size). To obtain a characterization of the "typical district," it would be necessary to weight the responses by the selection probability (found by taking the ratio of the district's enrollment used in the selection procedure to the total enrollment over all districts).

Our thinking at this stage was that the most desirable analyses would be those involving no weighting (except as may have been needed when dealing with very small district), and the sampling would reflect concern for representation of the primary target population -- teachers.

Revised Sampling Design

Toward the end of 1980, the initial sampling plan underwent a process of internal and external review. During the review process, it became clear that while the plan could be improved, certain features of the initial plan should be retained. For example, teachers remained as the primary focus of interest; MCT and geographical region were still used to define districts; low and high SES definitions of schools were as previously described in this section; the number of respondents of each type by district would remain constant (as seen in Table 5).

However, the review process revealed that while the PPS-based design would be adequate for the study's descriptive purposes, its capacity for allowing analytic, policy-relevant comparisons was limited. A series of project meetings led to the decision to replace the initial sampling procedure by a probability lattice methodology. The sample so produced, with minimal weightings, meets both the descriptive needs of the study -- to provide a nationwide picture of assessment practices and uses, and its analytic needs -- comparisons by SES, MCT, within districts, etc.

As was the case with the original design, in the revised plan 1,600 of the target population consisted of elementary and secondary teachers; approximately 400 were school principals, and another 120 were district testing officers.

The sampling was conducted in three stages:

(1) selection of 120 districts from a highly stratified sampling frame

(2) selection of two elementary and two secondary schools (size permitting) from each district

(3) selection of four teachers from each selected school

The selection of each stage was devised so that collectively the three stages produced a sample of teachers that was approximately self-weighting; that is, the overall selection probabilities for teachers was approximately equal.

First Stage (Selection of Districts)

Our sample called for a relatively small first-stage sample from a highly stratified sampling frame. With conventional stratification, the number of strata cannot exceed sample size, thus precluding its usefulness for our purposes. Although a number of stratification schemes introduced over the past thirty years do not have this limitation, most require symmetrical joint distributions over the stratification factors, which generally are not present in naturally occurring populations. Jessen (1970) has presented several schemes that do not require such symmetry under the collective label "probability lattice sampling."

With probability lattice sampling (PLS) (Jessen, 1970), we were able to obtain a sample that was similar to latin square experimental design. The sampling universe was stratified into several levels for each of several factors, and the sample that was selected simultaneously represented each level of each factor in predesignated proportions. This result is obtained with probabilities proportional to size even though the cells formed by the multiple stratification have different measures-of-size (MOS). Indeed, most of the cells in our sampling frame had zero MOS, (i.e., were empty).

Public school districts were the sampling units for the first stage. The sampling population excluded Alaska and Hawaii, as well as districts that are not unified. The data source was District File, a listing of all U.S. public school districts by Market Data Retrieval (MDR), 1980. The school districts in the sampling population numbered 13,815, with a combined reported student enrollment of 41,589,605.

Stratification

Five stratification variables were chosen to enhance the analytic and descriptive qualities of the sample:

(1) minimum competency testing status
(2) size of student enrollment
(3) SES of attendance area
(4) geographic region
(5) metropolitan status

Minimum Competency Testing Status. Districts were categorized according to the status of minimum competency testing (MCT) in their respective states. This categorization, based on Gorth (1980) and Kaufman (1979), reflects whether a MCT program exists, whether MCT is a requirement for promotion or graduation, and whether the state allows local districts the option of designing or selecting the tests to be used. Thus, there were five strata:

	# districts	% total enrollment
(1) MCT not required for graduation or promotion; no local option	2703	19

		# districts	% total enrollment
(2)	MCT not required for graduation or promotion; local options.	2065	13
(3)	MCT required for graduation or promotion; no local option.	980	18
(4)	MCT required for graduation or promotion; local options.	1778	16
(5)	No MCT program mandated for implementation by 1981 at state level.	6289	34
		13815	100

Size of student enrollment. The enrollment strata were designed to assure representation from very small, small, medium, large, and very large districts. In setting the class limits, special attention was paid to previous CSE research that found that the organization of district administration and use of resources in testing is significantly different for districts that are above certain size thresholds (Lyon, 1978). The five strata are:

		# districts	% total enrollment
(1)	- 4,999	12061	37
(2)	5,000 - 9,999	1059	18
(3)	10,000 - 24,999	514	18
(4)	25,000 - 44,999	105	8
(5)	45,000 -	76	19
		13815	100

Of these five strata, numbers (2) - (5) are identical with those used in the Lyon study just cited. An additional, smaller strata (1) was added here to assure representation of smaller districts.

SES of attendance area. The MDR data file indexes school districts into four categories based upon calculations of the Orshansky Index. These categories were collapsed into three strata:

	# districts	% total enrollment
(1) .1 - 4.9% (wealthiest)	1907	16
(2) 5.0 - 24.9%	9051	69
(3) 25.0 - (poorest)	2857	15
	13815	100

Geographic region. Four strata were defined in order to assure representation across the continental United States:

(1) Northeast -- Connecticut, Delaware, District of Columbia, Maine, Maryland, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont.

2718 districts 25% of total enrollment

(2) Southeast -- Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia.

1736 districts 24% of total enrollment

(3) Middle -- Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin.

5279 districts 27% of total enrollment

(4) West -- Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oklahoma, Oregon, Texas, Utah, Washington, Wyoming.

4092 districts 25% of total enrollment

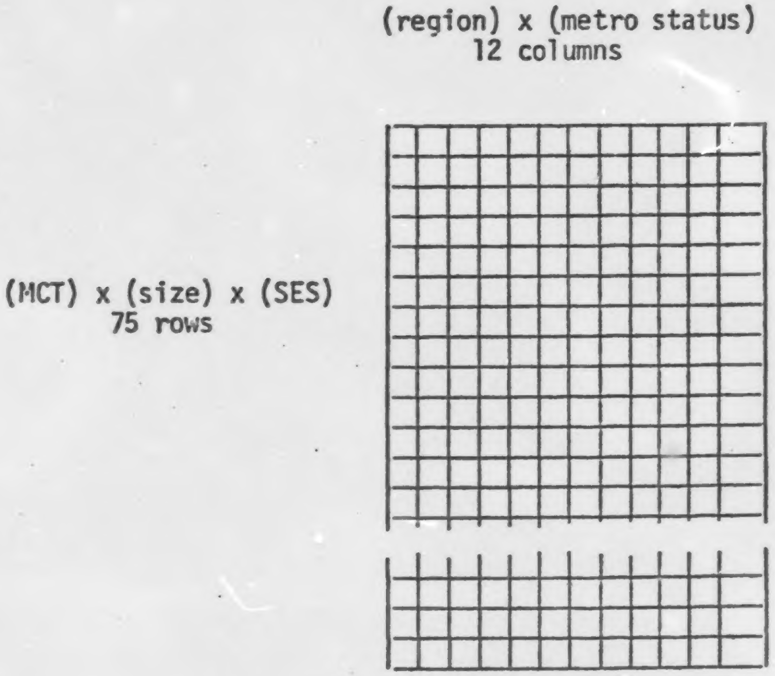
(These divisions are identical to those used by the National Education Association in its annual survey of teachers' activities and opinions.)

Metropolitan status. The MDR data file groups school districts into three levels of metropolitan status. These groupings were adopted as strata to reflect different degrees of urbanness:

	# districts	% total enrollment
(1) Central City	915	31
(2) Urban Fringe	3354	32
(3) Non-metropolitan	9546	37
	13815	100

Sampling Frame

From the five stratification variables, a 900-cell matrix was fashioned that has 75 rows and 12 columns:


(region) x (metro status)
12 columns
(MCT) x (size) x (SES)
75 rows

Upon allocating the 13,815 districts in the sampling population among the cells of this matrix, nine full rows were found to be unoccupied, as were an additional 436 cells. The occupied cells, then, numbered 356.

Sampling from this matrix in such a manner that each occupied row and column (but not each cell) is represented had the effect of crossing the two column factors in factorial fashion, and similarly crossing the three row factors in "near" factorial fashion; the exception, of course, is that nine combinations of the row factors did not exist in the population of districts. The methods used to achieve this sampling are described below.

Selection Procedure

The sample size of districts was set at 120 rather than 100 as envisioned in the original sampling plan. The larger sample size partly offset the fact that many of the smaller districts have only one high school, and some may have just one elementary school.

The selection probabilities for districts were set proportional to a measure-of-size (MOS), namely, student enrollments reported in the MDR data file. With properly coordinated selection probabilities at the successive two stages, our sampling of teachers theoretically could have been self-weighting (i.e., equal probability) without the inconvenience of a highly variable sample size for teachers. However, as amplified later in this section, analytic and cost considerations led us to modify the procedure so that the sampling in self-weighting was approximate rather than exact.

Weighting. To enhance the analytic characteristics of the sample we undersampled from two strata:

(1) MCT Stratum #5 -- no minimum competency testing program at state level.

(2) Size Stratum #1 -- districts with enrollment less than 5,000.

Undersampling from MCT Stratum #5 permitted the selection from the strata of correspondingly more districts that have greater analytic interest; i.e., those with some MCT program in force or to be implemented by 1981. A target of approximately 20 districts in MCT Stratum #5 was accordingly set, and the weight for this stratum was set at 0.6. A target of approximately 25 districts was set for Size Stratum #1 in order to avoid over-burdening the sample with small districts, which in some respects are of less interest to this study because they draw fewer federal and state dollars and allocate fewer local resources to testing (Lyon, et al, 1979). Accordingly, the weight for this stratum was set at 0.7. In order to accommodate these two weightings into the sampling frame, weights were required for other cells as well: cells that were jointly in MCT Stratum #5 and Size Stratum #1 were weighted by 0.42; cells that were part of neither of these strata were weighted by 1.50.

The implication of the above weighting scheme was to specify a sample of districts that was distributed across the various strata as depicted in Table 6. Table 6 also includes the actual sample allocation that resulted.

Cell selection. The first-stage selection of school districts was actually accomplished in two "sub-stages: -- 120 cell selections from

Table 6

Allocation of District Sample Among Strata

Variable	Level	Expected Allocation*	Actual Target Sample	Responding Districts**
MCT Status	1	27.1	27	22
	2	18.2	18	17
	3	28.5	28	21
	4	25.4	26	16
	5	20.9	21	15
Enrollment	1	25.5	25	19
	2	25.7	26	22
	3	26.7	27	22
	4	12.3	12	9
	5	29.8	30	19
SES	1	17.8	17	15
	2	85.4	86	61
	3	16.9	17	15
Region	1	31.4	31	22
	2	33.0	33	28
	3	27.5	27	22
	4	28.2	29	19
Metropolitan Status	1	47.4	47	33
	2	35.9	36	27
	3	36.7	37	31

---

*The fractional portion of the allocations should be interpreted as the probabilities with which an additional district should be selected from the respective stratum. For example, 27 districts are to be selected from MCT Stratum #1, with a ten percent chance of selecting a 28th district.

**Corrected weights corresponding to the figures in this column will be incorporated and used throughout the analyses of the final report. Preliminary results reported in this document were computed using equal weights.

the sampling frame, then the selection of one district for each cell selection.

The previously described PLS that we used required us to arbitrarily designate a "feasible set" of lattices for the frame, each of which satisfies the pre-designated quotas for each stratifying factor. For example, where the stratification quota is three, each feasible lattice designation is required to include three non-zero cells for that stratum. Each lattice was assigned a selection probability. The number of lattices in the feasible set and their selection probabilities are not jointly artibrary, but are determined by a set of decision rules that guarantee that the sum of probabilities for all lattices that include a particular cell is proportional to the MOS for that cell. Finally, observing the assigned selection probabilities, we selected one lattice from the feasible set to obtain the sample of cells.

Since some of the larger cells were designated more than once (in accordance with cell size), the total number of district cells in our selection lattice was only 98.

Selection of districts within cells. School districts were sampled from selected cells with probabilities proportional to MOS (again, district enrollment). The procedure was to list the districts within cells in alphabetic order, cumulate the MOS, then select a random number between one (1.0) and the total MOS for that cell. By matching the random number of the cumulative MOS, the sampled district was identified. This process was repeated in cells selected more than once.

Anticipating that some districts would refuse to participate in the study, a nonresponse strategy was developed. At the district level, non-cooperating districts were replaced from the same cells from which the refusals came.

Second Stage (Selection of Schools)

As noted earlier, the sample design called for selection with probabilities proportional to MOS of two elementary and two secondary schools from each selected district. (Many of the smaller districts, of course, yielded only one secondary school.) The procedure for this selection was as follows.

Before the initial district contact, a list of schools and their enrollments was obtained from data files prepared by the Office of Civil Rights and the National Center for Education Statistics. If the district was large, a pre-selection of eight elementary and four secondary schools was made using systematic sampling to select with probabilities proportional to size.

A protocol was then designed to structure initial telephone contacts with officials of selected districts. During the course of the district contact, the list of pre-selected schools was read to a district official, who was asked to rank them according to percent AFDC, percent receiving free lunch, or another locally salient poverty status variable. A cumulative list of the MOS for the ranked schools was calculated, and a systematic sampling was used to make the selections with probabilities proportional to MOS. In the case of large districts where a pre-selection was made as described above, actual selection of the four schools to be included in the sample was made with equal probabilities since a MOS selection had already been made. This selection process had the effect of stratifying within the district on the basis of poverty status. Where a school declined to cooperate (or the district refused to permit a sampled school to participate), another school in the appropriate poverty strata was chosen from the pre-selected subset.

Third Stage (Selection of Teachers)

The four teacher types (fourth grade, sixth grade, tenth grade language arts, tenth grade mathematics) were treated as separate populations. As noted earlier, a sample of two teachers was targeted from each type.

In order to complete the self-weighting nature of the survey sample, it would have been necessary to collect lists of the four teacher types for each school, or at least the numbers thereof. Selection probabilities would then have been calculated as functions of the MOS used in the second-stage selection of the respective schools:

$$\frac{2(K)}{(\text{MOS for school})}$$

where K, which is equal to 12 times the national average (or typical) pupil/teacher ratio, accounts for the fact that MOS in the first stages was based on student enrollments rather than numbers of teachers. Thus, the overall selection probabilities would have been constant except for the effect of the stratum weights applied in the first stage sampling:

$$\frac{100(W_i)(\text{district MOS})}{(\text{district MOS})} \times \frac{2(\text{school MOS})}{(\text{district MOS})} \times \frac{2(K)}{(\text{school MOS})} = \frac{400(W_i)(K)}{(\text{district MOS}}$$

where W_i is the stratum weight.

The above procedure would have provided the expectation of two teachers of each type from each school with just four distinct values for W_i. However, that procedure had two significant drawbacks: (1) it would have required additional phone contacts before sampling could be completed; and (2) the variation of sample size from each school could have allowed one or more respondents in a given category to be selected.

Because of these problems, we devised a procedure, based on previous CSE survey research, in which the respective principals were provided with a systematic sampling protocol enabling them to select two teachers at random from the appropriate categories. The variability in third-stage selection probabilities has slight if any effect, and the cost and time savings, as well as analytical advantages, are significant.

Principals were provided with extra questionnaires and directions for selecting, where possible, alternate respondents in the case of teacher non-response.

Questionnaire development. Questionnaire development drew upon the theory of test use and the conceptual scheme previously described. Development efforts were informed by the experiences of and information emanating from the various project planning activities. These sources enabled us to draw up specifications to describe content areas that the questionnaire items would tap. From these specifications, items sets were constructed for the teachers' and principals' questionnaires.

Draft questionnaires were reviewed by a variety of experts and practitioners as described in our interim report to the NIE, January 1981. In addition, questionnaires were piloted with principals and teachers (N=50) in three phases in both a large, urban school district and in a small, suburban district in the Southern California area. Each of the three pilot phases, in conjunction with expert reviews, provided information that was used to make successive revisions of each of the questionnaires. Teachers and principals participating in the field test were selectively interviewed about the instrument; all completed summary and item-by-item questionnaire review forms.

As a result of these procedures, all of the draft questionnaires were extensively modified, so as to focus more effectively on the information needed by the project while minimizing the response burden for individual respondents. The questionnaires were formatted so as to facilitate coding and data processing.

The teacher survey was constructed in two forms, elementary and secondary, to accomodate the obvious differences in class structure. At the elementary level, the questionaire asked teachers about testing in both mathematics _and_ reading. At the secondary level, the questionnaire asked mathematics and English/reading teachers about testing activities and test use _only_ in their subject field. Both teacher questionaires contain a common core of questions about perceptions, instructional organization, training and experience, leadership, and other contextual information. The principals' questionnaire is not differentiated by school level; it does differ, however, from the teacher questionnaire in the types of tests and the uses to which it refers so as to further reduce the response burden for teachers and to provide a means of confirming their content statements. Copies of the questionnaires in their final form are appended to this report.

Contacts with districts and schools. Following the revised sampling design, 120 school districts were selected during the first stage. (One hundred-fourteen school districts were actually selected. Since the probability of a given district's selection was proportional to a measure of size, several large districts were selected more than once. In these districts, the usual set of schools to be selected per district--two elementary schools and two high schools--were increased times the number of times the district was selected.)

Initially each sampled district was contacted by phone, and once the appropriate member of the district staff was identified, three matters were dealt with:

(a) The agreement of the district to participate was obtained. In many cases this was not granted until district personnel had had a chance to review the overall objectives of the project and the actual questionnaires to be used. This introduced some delay into the procedure, but the view was taken that the district was entitled to receive the fullest information about the project that the team could provide.

(b) The assistance of the district in selecting schools in accordance with the sampling frame was sought. Once schools were identified, procedures for obtaining the cooperation of the principals were discussed. In some cases, the district made all the arrangements and requested that CSE send all packets of questionnaires to the district for distribution. In the majority of districts, however, it was agreed that CSE would contact the schools directly.

(c) An indepth interview was conducted to determine the district's policy regarding assessment, mandated testing programs, inservice training for teachers in evaluation procedures, etc. Detailed information regarding the size, structure and demographic characteristics of the district were also recorded.

Although in some instances a single telephone call was sufficient, for most districts it was necessary to place a number of calls to several members of the local staff.

Questionnaires for elementary and secondary teachers and their principals were then sent to the selected districts in the following manner. A package containing the principal's and teachers' questionnaires was sent to the principal of each selected school (or as noted above, to the district headquarters). The principal, in addition to completing and returning his/her own questionnaire, also distributed the questionnaires to the teachers following the method described in the protocols. Teachers then completed and returned their questionnaires to CSE.

Returns from the first round of the survey were disappointingly low (see Table 7). Problems in printing the questionnaires and in getting approval for some districts to participate in the study, combined with unanticipated delays in the mail and at district offices meant that a substantial number of the sampled schools received the packets of questionnaire material very close to the end of the school year. In some schools, the questionnaires did not get into the hands of the selected teachers before the beginning of the summer vacation, and other schools indicated to CSE that they would be unwilling to provide the requested information because of the pressure of other work.

Table 7

Response Rates

	Target Sample	Responses received by July 31	(Percent of target)	Responses received by Nov. 6	(Percent of target)
Districts responding	114	53	(46%)	91	(80%)
Principal's Questionnaires	400	144	(36%)	222	(56%)
Secondary Teachers Questionnaires	800	244	(31%)	372	(47%)
Elementary Teachers Questionnaires	800	305	(38%)	488	(61%)

It was felt that these factors accounted for the initially low response rates. However, this same low rate jeopardized the validity of the sampling and hence the credibility of the results, and so it was decided to extend the deadline for the return of questionnaires into the fall. Schools that did not respond in the May/June period were contacted and encouraged to send back completed questionnaires in September. Replacement questionnaires were mailed to those schools who reported having mislaid, or not having received the original packages. In addition, the opportunity was taken to substitute districts and individual schools that communicated to CSE their decision not to participate in the project. These were replaced with others drawn from the original master list and using the same sampling procedures. Thus, the survey was essentially reactivated at the end of August and during September, and a major effort was made to expand the pool of data that would be available for analysis. Even so, schools proved slow to respond and questionnaires were still being returned at the end of October.

To obtain the most comprehensive and reliable picture of the national scene with regard to test use, it was considered important to include as many completed questionnaires as possible in the main analysis. For this reason, the deadline for the return of questionnaires to be included in the statistical analysis was delayed until November 6. Table 7 gives the number of completed questionnaires received by that date, and that constitute the sample for the main analysis. The response rates, even at the cut-off date, were still comparatively low, but since a certain shrinkage had been anticipated in drawing up the sample design, it was considered that the number of returns was sufficient to permit the original analysis to proceed.

The uneven pattern of response, however, casts some doubt on the validity of the achieved sample and certainly ensured that the intended simple weighting design outlined in an earlier section of this report would not be adequate. Unfortunately, the checking of the validity of the sample and the calculation of the weights to be used in the main analysis, could not be carried out before the cut-off date for receipt of questionnaires, and so the final analyses have not yet been completed. However, in order to test out the file handling, data editing, and stitistical procedures, and to give an indication of the major trends in the data, most of the intended analyses have been run on the subset of responses that were received by July 31. Selected summaries of these are presented in the following pages, but it must be noted that these initial analyses were carried out using unit weights for each district so they should be regarded as illustrative rather than as representative of the pattern of results to be expected in the full report.

Analytic framework. It will be recalled from our 1980 report to the NIE that the project's domains of interest (as depicted on page 17 in the present report) and data collected in the project's planning stage generated a series of hypotheses or questions to be explored in the national survey. These areas are summarized here to serve as background for the discussion of survey findings.

Federal/state/local testing requirements influence the distribution and frequency of types of testing at local sites, and thus bear upon patterns of test use. Testing interventions such as minimum competency testing, therefore, may impact on the organization of curriculum and instruction.

The organization of curriculum and instruction constitutes a major influence on the nature of teachers' routine, practical activities and decisions. We hypothesized, therefore, that a greater variety and number of available instructional alternatives in the classroom and school will increase the routine tasks and decisions that require assessment information, and so influence both the patterns of testing that occur locally and the ways test scores are used locally.

The nature of teachers' routine, practical activities and decisions is assumed to vary with the types of students enrolled in the school and assigned to a teacher's classroom. Thus, the types of tests given locally and the uses of test results are likely to vary with the demographic and achievement characteristics of students in the school and classroom.

As teachers go about the accomplishment of their practical tasks and decisions, the instances in which they refer to test scores and the ways in which they "weight" test scores are assumed to vary with their perceptions (opinions, values, understandings) of tests and types of tests.

As teachers assess particular tests' strengths and weaknesses and their appropriate uses, they will draw upon their educational and practical experiences with respect to testing. Thus, their training and experience are likely to bear ultimately on their practical decisions about which type of test scores to use and how to use them.

We assume that innovative district and school leadership can provide inservice training experiences that change teachers' perceptions of the utility of particular tests and types of tests, thus influencing teachers' practical test-use decisions. District and school leadership can also act to generate tests, testing programs, and practices that facilitate teachers' accomplishment of their routine tasks under the practical exigencies of their environments.

The types of tests given locally, and the purposes for and frequency with which they are given, will influence local types of test score use. The presence/absence of one type of test may influence the use of scores from another type. For example, the use of minimum competency tests for graduation may encourage teachers to use the results of other kinds of tests to measure students' progress toward the attainment of minimum competencies.

THE ELEMENTARY SCHOOL TEACHER SAMPLE

This section of the report provides some background information on the elementary school teacher sample and some of the characteristics of their classrooms. It also discusses the degree to which teachers make use of various resources, the kinds of assistance on matters of assessment they receive from their districts, school or district level training they receive in testing and assessment, district use of assessment vis-a-vis teachers' instructional practice, and patterns of district reporting of test results back to teachers. This information is offered as precursor to a subsequent section dealing with teacher attitude toward testing and their reported uses of assessment results.

Teachers' Professional Background

The first section of the elementary school teachers' questionnaire asked respondents a series of questions about their professional background. The first of these questions dealt with the number of years the teacher had been teaching. Table 8 below illustrates the responses to this question, with years of teaching primarily broken down into five year periods.

Table 8

Years of Teaching
(N-304)

Number of Years Teaching	Number of Respondents	Percentage
1 - 5	54	17
6 - 10	83	28
11 - 15	71	23
16 - 20	42	14
21 - 25	27	9
26 - 30	20	6
31 - 40	7	3
	304	100

This question was followed by an item asking respondents how many years they had been teaching in their present district. Table 9 below indicates patterns of teacher responses to this item, using the same breakdown of years as followed in the preceding table.

Table 9

Number of Years Teaching in the Present District
(N-304)

Number of Years in District	Number of Respondents	Percentage
1 - 5	98	32
6 - 10	72	24
11 - 15	76	25
16 - 20	30	10
21 - 25	12	4
26 - 30	11	3
31 - 40	5	2
	304	100

These data indicate a certain amount of stability among the elementary teacher population. Beyond the 32 percent who have been in their district for one to five years, an additional 25 percent have been in their district for six to ten years and 25 percent for eleven to fifteen years. An additional 10 percent have been in their present district for 16 to 20 years, with the remaining 9 percent serving in the same district in excess of 20 years.

The next questionnaire item asked teachers the highest diploma or degree they have received. Of the 298 respondents, 178 (58%) had received a bachelors as highest degree with the remaining 125 (42%) reporting a masters as highest degree; none of the respondents indicated receiving a doctorate.

The year that respondents received their degree is indicated in Table 10 below, which again breaks down year of degree in five-year periods.

Table 10

Year Degree Was Received
(N = 297)

Years	Number of Respondents	Percentage
1935 - 41	7	3
1946 - 50	8	3
1951 - 55	12	4
1956 - 60	20	7
1961 - 65	33	11
1966 - 70	49	17
1971 - 75	79	26
1976 - 81	89	30
	297	100

These data would suggest that the elementary teachers constitute a fairly youthful population, with 56 percent having received their highest degree in the last 10 years, and with almost 75 percent having received their highest degree in the last 15 years.

Two-hundred thirty-six teachers indicated that they had received additional credits/units beyond their last degree, with a median value of 23, and 16 teachers reporting 100 or more.

Classroom Characteristics

Of the 305 teachers responding to the questions on the number of grades in their regular classroom, 273 (90%) indicated that they teach only one grade; 22 (7%) that they have two grades; 7 (2%) that they teach three grades; 2 (1%) that they have four grades; and 1 teacher reported having five grades. The following picture, as indicated in Table 11 , reflects the "modal" grades taught by the responding sample.

Table 11

"Modal" Grades Taught
(N = 305)

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7
No. of Teachers	5	5	11	134	43	104	2
Percent	2%	2%	4%	44%	14%	34%	1%

Of the total sample, then, 281 teachers, or 92%, clustered around grades four, five, and six--the targeted grade levels of CSE's national survey.

The teachers were also asked a question on the average numbers of students they presently have in their classrooms. Table 12 below indicates the patterns of average numbers of students.

Table 12

Average Number of Students in a Classroom
(N = 302)

Number of Students	Number of Teachers With This Size Class	Percentage
(up to) 15	3	1
16 - 20	17	6
21 - 25	76	25
26 - 30	138	45
31 - 35	55	19
36 - 41	7	2
41 plus	6	2

It would seem that the "average" teacher has a class consisting of 26 to 30 students, and that the great majority of the teachers have classes comprised of 21 to 30 students.

The teachers also indicated on the survey their current teaching responsibilities in reading and math. Of the 301 respondents to this item, 247 teachers (82%) teach both reading and math; 14 teachers (5%)

indicated that they teach math only; 36 teachers (12%) that they teach reading only; and 4 teachers (1%) that they presently teach neither. Most teachers devote four to seven hours a week on reading, and four to six hours a week on math. In terms of the different curricular levels at which they must teach in a given classroom, most teachers reported having students representing three to five different reading levels; in math, however, most teachers reported having only one to three different student levels.

Use of Resources

The next item on the survey asked teachers a variety of questions dealing with specific resources that they may use in the classroom. Response rates for these questions ranged from 260 to 290. The following picture emerged from the teachers' responses about their use of these resources. In terms of teachers having another adult under their supervision to help with small group/individual student work, almost 60% indicated that this resource is not available to them either for reading or math. Another 10% indicated that such a resource is available, for both reading and math, but is not used, while 2%, for both reading and math, indicated that an adult is available but used very infrequently. A few teachers (1 to 3%) indicated for reading and math that an adult might be used once or twice a month, while 20 to 25 percent indicated, again for reading and math, that an adult aide might be used once or twice a week.

Another item in this series on the survey asked teachers if they can divide up students for extra help among other teachers. Forty-five percent of the teachers indicated they do not have this resource in reading, and approximately 55 percent that they do not have it for math. Nine percent reported that the resource is available in reading but not used, and 12 percent that it is available in math but not used. Again, a few teachers reported using additional teachers a few times a year in reading and math help; and about 34 percent reported this practice once or twice a week in reading, and 25 percent once or twice a week in math.

In terms of the availability of instructional machines, such as audiovisuals, computer terminals, etc., for students' independent work, approximately 35 percent of the teachers reported that they are not available in either reading or math; another 10 percent reported that such technology, though available, is not used in either reading or math. The remainder of the teachers reported that they use instructional technology to varying degrees; for both reading and math, this use ranges from once a year, to several times a year, to once or twice a month, to a few times a week. Each of these categories of use accounts for approximately 5 to 10 percent of the population for both reading and math.

Similar patterns of teachers' response, for reading and math, emerged for such resource possibilities as working with other teachers for planning and developing tests and other evaluation assignments, and for the availability of specialists to whom students can be sent for special work; however, in the case of specialists, many more teachers report frequent use of this resource, for both reading and math, than is the case in some of the other resources discussed above.

Most of the teachers reported that alternate published or teacher-made materials are available and quite frequently used for students' special needs.

In terms of the three remaining resources queried on the survey--having someone available to read and/or grade tests and other student assignments; quick, computerized scoring and analysis of tests; and "item banks" to draw upon in making up teacher tests--a clear picture emerges; these resources are simply not available to the vast majority of the respondents in either reading or math (the negative response rate range from approximately 65 to 75 percent of the respondents).

District Assistance

The next item asked about the district provision of help to teachers in matters related to student assessment. Approximately 300 teachers responded to the questions associated with this item.

In terms of receiving help in the administration of required tests, 248 teachers (82%) indicated that such help is available; for 18 percent of the teachers it is not available. Of the teachers receiving this kind of help, most indicated that it is relevant or very relevant to their specific classroom tasks.

Two hundred fifty-six of the teachers (84%) receive assistance in analysis and explanation of test results; the remaining 16 percent do not. Of those receiving this help, again most of them noted that it is relevant to their classroom work.

The picture reflecting teachers receiving assistance in alternative ways (other than tests) to assess student achievement is quite clear-cut; 50 percent of the sample receive this kind of assistance and the other 50 percent do not. Of those receiving this help, most feel that it is relevant to their classroom responsibilities.

A similar picture emerged with respect to preparing students for particular kinds of tests; a little more than 50 percent of the teachers do not receive this help, and a little less than 50 percent do. Of those teachers who do, again they find it relevant.

Almost 60 percent of the teachers report receiving assistance in interpreting and using the results of different kinds of tests; the remaining 40 percent do not. Again, most teachers who receive this assistance find it useful.

About half the respondents receive help in ways to tie their teaching content to that of required tests and half do not. Again, those receiving this assistance find it useful and relevant to their classroom work.

In terms of help in constructing or selecting good tests, the vast majority (approximately 85%) do not receive this kind of assistance; of those who do, most find it relevant or very relevant to their classroom work. Similarly, most of the teachers (65%) do not receive training in the use of assessment results to improve the instructional program, but those who do find it relevant.

District Training/College Courses

Of the 100 teachers responding to the next item on the survey, approximately 50 percent indicated that in the last two years they have attended one to five hours of meetings on the topic of selecting or constructing tests or establishing district testing policies. Another 25 percent noted that they have attended six to ten hours of such meetings. These data tend to corroborate the item discussed immediately above. That is, only 100 of our teachers receive district training in constructing or selecting good tests, and of these, training has amounted to only three or four hours for 50 percent of the recipients and may have dealt as much with testing policy as with test selection/construction. In terms of district inservice on other topics related to student assessment, most of the 183 respondents to this item indicated such inservice in the range of one to ten hours in the past two years.

Sixty-six teachers reported taking college courses in the last two years that were devoted exclusively to student assessment. Of these teachers, about 30 percent have taken two to five hours; 20 percent six to ten hours; 10 percent 12 to 15 hours; 6 percent have taken 16 to 20 hours; 14 percent 30 to 35 hours; and 4 percent 35 to 40 hours. The remaining 15 percent or so reported taking college courses on assessment in excess of 45 hours.

District Uses of Assessment Information

The next item on the survey asked a series of questions about school administration uses of assessment information vis-a-vis teachers' instructional practices. Approximately 300 elementary teachers responded to this item. In terms of school administration review of test scores with teachers for the purpose of identifying curriculum content areas needing extra emphasis, 35 percent of the respondents indicated that such practice is a regular occurrence and part of the school's routine procedure, while another 25 percent indicated that it happens quite frequently but not on a routine or regular basis. For the remainder, it happens rarely (28%) or not at all (14%).

For almost 30 percent of the teachers, a school administrator observes the teachers, reviews his/her instructional plans, etc., on a regular/routine basis to make sure that students' needs as indicated by test scores are emphasized; for another 25 percent this happens quite often but is not regular or routine. Of the remainder, about half reported that such practice happens rarely and half that it does not happen at all.

In terms of teachers being required to turn in the scores or grades of tests or assignments that they routinely give, about 15 percent of the respondents indicated this happens regularly and routinely; for another 5 percent, this practice goes on quite often but not on a routine basis. For about 15 percent this happens rarely, and for the remaining 65 percent it does not happen at all.

The final question in this series asked teachers whether their school evaluates their teaching on the basis of students' test scores, and/or establishes test score goals for the students and the teacher to meet. For the vast majority of the population (70%) this practice is not followed. Of the remainder, about 17 percent indicated that it happens rarely; about 6 percent each reported this practice as either regular and routine, or frequent but neither regular nor routine.

District Reporting of Test Results

The last of the background questions asked teachers a series of questions dealing with test turn-around time, usefulness of test reporting formats, and encouragement to teach in the basic skills.

Of the 300 or so teachers who responded, 133 (44%) indicated that they receive test results from the district soon enough that they can use the results for instructional modification; another 139 (46%) noted that they receive the results too slowly to be of use in modifying teaching; the remaining 10 percent indicated that the question does not apply.

In terms of the district reporting test results in a way that enables the teacher to use them, the vast majority (72%) indicated that they receive results that are detailed and in a useful format (though perhaps they arrive too late for this potential to be realized, as was suggested immediately above). Another 21 percent answered that little useful information is provided in the way of reported test scores; the remainder indicated that the question does not apply.

In connection with an assessment program and district encouragement of teacher emphasis on the teaching of basic skills, 95 percent of the respondents indicated that their districts do follow this practice.

Teacher Attitude Toward Tests and Test-Related Issues

Approximately 300 elementary school teachers responded to a series of items probing teacher attitudes toward tests and test-related issues. Table 13 below illustrates the more prevalent trends emerging from these items.

Table 13

Teacher Attitude Toward Tests and Test-Related Issues
(N = 300)

Item	Percentage of Teachers in Agreement
Testing motivates my students to study harder.	60
Commercial tests are usually of high quality.	60
The content (or skills) on most required tests is very similar to the content (or skills) that I teach.	75
The pressure that testing exerts on the schools has a generally beneficial effect.	45
Recently, I have been spending more teaching time preparing my students to take required tests.	50

Table 13
(continued)

Item	Percentage of Teachers in Agreement
The tests developed in our district are very good.	60
The curriculum today demands more complex student thinking than in the past.	70
Teachers should not be held accountable for students' scores on standardized achievement tests or tests of minimum competency.	70
In our school, students are more rigidly tracked than they were two or three years ago.	55
Tests of minimum competency/proficiency/functional literacy should be required of all students for promotion at certain grade levels or for high school graduation.	80
Tests of minimum competency are frequently unfair to particular students.	55
As a result of minimum competency tests (and similar programs), parents are contacting schools about their children more frequently or in greater numbers.	55
Tests of minimum competency have affected (would affect) the amount of time I can spend teaching subjects or skills that the tests do not cover.	60
In our school, testing programs are generally held to be much less important than the social problems with which we are concerned.	40
Basic skills teaching (including remedial work) is now consuming a substantially increased proportion of our school's educational resoruces.	90
The proportion of our school's resources now allocated to basic skills teaching is so great as to detract from the quality of our total educational program.	20

From the above patterns, it appears that teachers see some beneficial effects accruing from testing and test-related matters, and that the tests they speak about are frequently seen to be of generally high quality and match what they teach. Many of the teachers see the importance of minimum competency tests, although more than half of our respondents have reservations about the fairness of such tests for certain kinds of students. Perhaps as a function of minimum competency tests, many teachers report that their students are more rigidly tracked than was the case in the recent past, which might concern the majority who believe that today's curriculum is more complex and demanding upon the student than was previously the case.

Basic skills teaching appears to consume an increasing proportion of school resources for the majority of teachers, and affects the amount of time, for more than half of the sample, that they can devote to other subjects. More than half the teachers state that their testing programs are held to be more important than the social problems with which they are concerned. However, the majority do not believe that the proportion of school resources given over to basic skills is so high that it detracts from the quality of the total educational program.

Teacher Uses of Assessment Results

Table 14 following provides a summary of the elementary school teachers' responses to a series of questions on their use of various kinds of information for specific decision-making purposes. These decision areas were concerned with the importance of different kinds of information for: (1) planning teaching at the beginning of the school year; (2) for initial grouping or placement of students for instruction;

Table 14

Teacher Use of Assessment Information for Different Decision-Making Purposes

(Percentages reporting use of this information for the specified purpose)

Source/Kind of Information	Planning Teaching at Beginning of School Year		Initial Grouping of Students		Changing a Student from One Group or Curriculum to Another		Deciding on Students' Report Card Grades	
	Reading	Math	Reading	Math	Reading	Math	Reading	Math
Previous teacher's comments, reports, grades	61	55	62	56				
Students' standardized test scores	57	56	55	53	54	52	19	17
Students' scores on district continuum on minimum competency tests	53	50	52	50	46	42	22	20
My previous teaching experience	96	95						
Results of tests included with curriculum being used			77	69	84	81	72	73
Results of other special placement tests			62	60				
Results of special tests developed or chosen by my school					54	48	38	39
Results of tests I make up			83	88	81	85	93	96
My own observations and students' classroom work			97	98	99	100	98	98

(3) for making decisions to change a student from one group or curriculum to another, or to provide remedial or accelerated instruction; and (4) for making decisions on students' report card grades. The data appearing in the table indicate the percentage of teachers who rated a given information source as crucial or important. Response rates ranged from 260 to 300, approximately.

Several conclusions seem to be warranted, at least tentatively, on the basis of these data. For example, whether a respondent is describing assessment information use for reading or math, the relative weight teachers ascribe to a given kind of information remains fairly constant in the decision-making process.

In terms of decisions about planning for instruction, it is clear that the individual teacher's previous classroom experience is by far the single most important kind of information. Students' scores on standardized and other formal tests, however, appear to be almost as important in this decision as comments and other information about students offered by their previous teachers. This finding confirms conclusions drawn from previous CSE work on test use. It is interesting to note, however, that for a sizeable number of teachers, a number that is sometimes in excess of 50 percent of the sample, students' scores on standardized and other formal tests are important not only for initial placement decisions (also found in previous CSE data) but also for decisions about changing a student from one group to another or one curriculum to another. That is, for a sizeable number of elementary school teachers, formal test scores assume importance not only at the beginning of the school year but also during the school year. This conclusion does not run counter to previous CSE findings, because information is used in conjunction with other kinds of data in the teachers' decision-making--again a finding supported in previous CSE data. Further, in terms of decisions about initial placement, by far the most important kind of information is teacher observations and students' classroom work, followed by the results of tests teachers have made up themselves and the results of tests that come with the curriculum they use.

An almost identical pattern appears for decisions about grouping and/or instructional changes for a student and for decisions about students' report card grades, with the exception that for these last two decisions, the weights teachers ascribe to student scores on standardized and district continuum or competency tests fall off drastically.

As we have reported previously, teachers appear to rely on multiple sources of information for making their classroom decisions. The use of "formal" tests is more dominant early in the school year, and as the year advances and different kinds of decisions about individual students, groups, and classes have to be made, teachers seem to switch more to use of their own professional experiences, observations, students' classroom work, the results of teacher-made tests, and tests that come with the curriculum informing their teaching.

One final observation should be made about these data on teacher use of assessment information. The percentages shown in Table 14 above reflect numbers of teachers for whom an information source is crucial or important for a given kind of decision. The percentages not accounted for in these data constitute numbers of teachers rating a given kind of information as slightly important or unimportant. In those cases where percentages of teachers reporting an information source as important lie in the 50 to 60 percent range, and therefore 40 to 50 percent of the sample are not accounted for, generally about another 25 percent of the teachers find the information to be at least slightly important. Exceptions to this pattern, of course, are students' scores on standardized and district continuum or competency tests in making decisions about students' grades, where anywhere between 35 to 50 percent of the teachers find these kinds of information as unimportant.

THE SECONDARY SCHOOL TEACHER SAMPLE

Before presenting the preliminary findings on secondary teachers' attitudes toward and uses of assessment information, we will again offer some relevant background information on the characteristics of this population, as well as on testing and test-related matters in their schools and districts.

Teachers' Professional Background

Table 15 below presents, for English and mathematics teachers, the number of years they have been teaching, broken down into five-year segments.

Table 15:
Years of Teaching

Number of Years Teaching	English (N=124) Number of Respondents	English (N=124) Percent	Mathematics (N=117) Number of Respondents	Mathematics (N=117) Percent
1 - 5	17	13	18	15
6 - 10	37	30	28	24
11 - 15	35	27	40	33
16 - 20	17	15	18	16
21 - 25	12	10	7	6
26 - 30	5	5	3	3
31	1	1	3	3
	124	100	117	100

The next item on the survey asked teachers how long they had been teaching in their districts. Table 16 below shows the response patterns to this item, again in five-year periods.

Table 16 :

Number of Years Teaching in the Present District

Number of Years in District	English (N=123) Number of Respondents	English (N=123) Percent	Mathematics (N=117) Number of Respondents	Mathematics (N=117) Percent
1 - 5	39	32	36	30
6 - 10	33	27	25	21
11 - 15	32	27	34	29
16 - 20	14	11	15	13
21 - 25	4	3	3	3
26 - 30	1	1	1	1
31 - 35	0	0	3	3
	123	100	117	100

These percentages by years of service in the district are roughly the same as those found for the elementary school teachers, and indicate a similar degree of stability among the two samples.

Responses to the question on the highest diploma or degree received by the secondary teachers were as follows. Of the English teachers, 53 respondents (43%) list a bachelors as highest degree received, 67 (54%) a masters, and three teachers (2%) have received a doctorate. The mathematics teachers report that 56 (48%) have a bachelors as highest degree, 60 (51%) a masters, with one math teacher having obtained a doctorate.

The year that the English and math teachers received their degrees is indicated in Table 17 below.

Table 17 :
Year Degree Was Received

	English (N=124)		Mathematics (N=115)	
Years	Number of Respondents	Percent	Number of Respondents	Percent
1940-45	3	3	1	1
1946-50	1	1	3	3
1951-55	5	4	1	1
1956-60	8	7	7	7
1961-65	14	11	17	15
1966-70	26	20	25	21
1971-75	35	29	30	25
1976-81	32	26	31	28
	124	100	115	100

These data are again similar to those provided by the elementary school teachers; the secondary teachers show almost identical patterns of "youthfulness" or time-in-teaching.

Ninety-six of the English teachers and 91 of the math teachers reported that they have received additional credits beyond their last degree. Both samples show a median value of 24, with one teacher in each population reporting 100 or more extra credits or units received.

Classroom Characteristics

Approximately 120 respondents each from the English teachers and the math teachers answered a series of questions concerning their classroom characteristics. Among these characteristics were numbers of grades in their class, the grade in which teachers have the greatest numbers of students, the average number of students they presently have in their classrooms, their teaching responsibilities in English and math, numbers of hours of instruction they provide in these subjects, and the range of curricular levels at which their students are working.

For the English teachers, 30 respondents (24%) indicated that they teach only one grade: 42 (34%) that they teach two grades; 34 (27%) that they presently teach three grades; and 18 (15%) that they teach four grades. For the math teachers, 16 respondents (14%) indicated that they presently teach only one grade; 27 (23%) that they have two grades; 29 (25%) that they have three grades; and 41 respondents (35%) that they teach four grades. The "modal" grades taught by these teachers, expressed as a function of the grades in which they teach the greatest number of students, appear in Table 18 below.

Table 18:
"Modal" Grades Taught

	Grade 9		Grade 10		Grade 11		Grade 12	
	English	Math	English	Math	English	Math	English	Math
No. of Teachers	2	0	116	113	4	2	1	0
Percent	2	0	94	97	3	2	1	0

For the total sample of secondary teachers, then, both English and math, approximately 95% cluster at the tenth grade, the target grade of the national survey.

Table 19 below shows the average numbers of students in the English and math teachers' classrooms. Approximately 115 teachers responded from each sample.

Table 19:

Average Number of Students in a Classroom

	English Teachers		Mathematics Teachers	
Number of Students	No. of Teachers With This Size Class	Percent	No. of Teachers With This Size Class	Percent
10 - 15	4	4	8	8
16 - 20	15	13	10	9
21 - 25	38	33	33	30
26 - 30	40	35	41	37
31 - 35	14	12	9	9
36 - 40	1	1	1	1
41 plus	2	2	7	6

As was the case with the elementary teachers responding to the survey, it appears that the "average" secondary teacher, whether in English or math, teaches a class consisting of 26 to 30 students, that the vast majority teach classes consisting of 21 to 30 students, and that an additional 10 percent or so teach classes having 31 to 35 students.

In terms of the subjects they teach in their tenth grade classes, 123 English teachers (99%) and 116 math teachers (99%) teach English or math only; one respondent teaches both English and math in the tenth grade. Most of the English teachers and most of the math teachers report that students in each of their classes receive four to six hours of instruction/classwork each week in the subject area.

In the matter of the range of curricular levels at which they must teach in their subjects, different patterns appear for the English and math teachers. For the English teachers, approximately 30 percent teach at only one level and 35 percent have two levels; another 27 percent teach three different levels.

The remaining seven or eight percent have four or five levels. For the math teachers, approximately 57 percent teach at only one level, 23 percent at two levels, and 15 percent at three levels; three percent of the math teachers have four different levels, and two percent reported having more than five curricular levels.

Use of Resources

The secondary teachers' responses (N= approximately 120 for each sample) to questions dealing with resource availability and use indicate that certain kinds of resources are not available to most tenth-grade teachers, whether they teach English or math. There are also some resources which apparently are available to most tenth-grade teachers, again regardless of whether they teach English or math. There are one or two resources which are available to around half of the tenth-grade teachers regardless of subject taught, and one or two resources for which patterns of availability appear to differ as a function of subject taught.

One of the resource options queried on the survey dealt with the availability of another adult under the teacher's supervision to help with small group or individual student work. Approximately 80 percent of the English teachers and 85 percent of the math teachers report that this resource is not available; an additional five percent for each sample report the resource is available but not used. The 15 percent of the English teachers who do have and use this option vary in degree of use from once or twice a year, once or twice a month, to once or twice a week. Of the ten percent of the math teachers who do have and use this option, most frequently cited levels of use are once or twice a year and once or twice a week.

For about 70 to 75 percent of both the English teachers and the math teachers responding, the resource option of dividing their students among other teachers for extra help also appears to be unavailable. In both samples an additional ten percent report that this option is available but not used. For those few teachers who report having and using the option, degree of use varies from once or twice a year, to once or twice a month, to once or twice a week.

Having someone to help the teacher with reading, correcting, or grading the tests or other assignments they give to students does not appear to be available to most tenth-grade teachers. Approximately 70 percent of both the English and math teachers report the option as unavailable, and another five to ten percent report the option is available but not used. Of the remaining 20 percent or so who do have and use this resource, degree of use varies, but highest levels of use reported are once or twice a week.

There are one or two resources, on the other hand, which do appear to be available to most tenth grade teachers. For instance, approximately 85 percent each of the English and the math teachers report that alternate published or teacher-made curriculum materials are available to meet students' special needs. Almost all of the English teachers use this resource, with almost half of them reporting weekly use. Of the math teachers, about 12 percent do not use this option, and the remaining 70 percent or so report most frequently using the option several times a year or at the weekly level.

About 75 percent of the English teachers and 80 percent of the math teachers report that they have the option of working with other teachers to plan and develop tests or other assignments; about 10 percent of each sample report that they do not make use of this resource. English teachers' most frequently cited levels of use are once or twice a year and about once a week. Math teachers' most frequently cited degrees of use are about the same as those of the English teachers, with the exception that more math than English teachers report use of this resource on a weekly level.

For about half of the tenth grade teachers, in both subject areas, quick, computerized scoring and analysis of tests is reported as being available. But about 20 percent of the English teachers and 15 percent of the math teachers report that they do not use this option. Of the teachers who do use the option, most English and math teachers report degree of use at a few times a year; about five percent of each sample report using the resource once or twice a month, and a few teachers report use at once a week.

The numbers of teachers for whom "item banks" of test questions are available which they can draw upon for making up their own items are roughly similar to the numbers reporting for quick scoring of tests. This resource is available for about half of the math teachers and for almost half (46%) of the English teachers. About seven percent of each sample report they do not make use of this resource. Degree of use varies across both populations from once or twice a year, to once or twice a month, to once or twice a week.

The availability of the two remaining resource options queried in the survey appears to fluctuate somewhat more than those reported above in terms of the subject taught. For example, while only about 40 percent of the math teachers report that there are specialists outside their classroom to whom they can send students for special help, this option was reported as available by about 55 percent of the English teachers. About 10 percent of each population report they do not make use of this option; for those who do, degree of use varies from once or twice a year -- the most frequent response -- to once or twice a month and once or twice a week.

The availability of instructional machines, such as audiovisual equipment and computer terminals for students' independent work, also varies by subject taught. However, for this resource, it is only available for about half of the math teachers, while for the English teachers it is reported as available by about 65 percent of the sample. About 15 percent of the math teachers report they do not use this option, while less than 10 percent of the English teachers report they do not use it. For the math teachers, most frequently cited degrees of use are once or twice a year and once a twice a week; for the English teachers, the most frequently cited degrees of use are once or twice a year, once or twice a month, and once or twice a week.

District Assistance

The next part of the survey asked the secondary teachers a series of questions about the kinds of assistance their districts provide in matters related to student assessment. Some clear patterns emerge from the responses of the approximately 120 English and 120 math teachers who responded.

When we look at the responses of all the secondary teachers responding, in only one area of district assistance in assessment do the majority of teachers in both samples indicate receiving help; that help is in the matter of the district providing analysis and explanation of state, district, or school test results. For this item, 71 percent of the English teachers and 57 percent of the math teachers responded that this kind of help is provided. Of the 71 percent of the English teachers who indicated that they do receive this help, 53 percent noted that it is relevant or very relevant to their specific classroom tasks; 15 percent that it is slightly relevant; and only three percent that the help is not relevant. Of the 57 percent of the math teachers receiving this kind of help, about 40 percent indicated that it is relevant or very relevant to their classroom work;

15 percent that it is at least slightly relevant; and one teacher that the help is not relevant.

District provision of assistance to teachers also seems to occur for the administration of tests required by state, district, and/or school, but more so for English than math teachers. Sixty-three percent of the English teachers indicated that such help is available to them, but only 42 percent of the math teachers noted that it is available. Of the English teachers, about 47 percent responded that the help is relevant or very relevant to their classroom work; about 13 percent that it is slightly relevant; and two percent that it is not relevant. Of the 42 percent of the math teachers who do receive this district help, about 27 percent responded that it is relevant or very relevant; the remaining 15 percent are almost equally divided between slightly relevant and not relevant.

The two areas above are the only ones for which districts consistently make an effort, at least as perceived by the teachers, to provide assistance in matters related to student assessment. For the remaining six items querying district assistance the pattern is clear; most teachers, in both English and math, report that the assistance simply is not available.

For example, when asked if their districts provide help in selecting or constructing good tests, 80 percent of the English teachers and 85 percent of the math teachers reported that their districts do not. For those English and math teachers who do receive this assistance, most reported that it is relevant or very relevant to them; only about three percent of each sample indicated that it is only slightly relevant.

In the area of district help in alternate ways (other than tests) that teachers can use to assess student achievement, 68 percent of the English teachers and 78 percent of the math teachers reported that it is not available.

Again, of those teachers receiving this help, most indicated that it is relevant or very relevant, and about six or seven percent that it is only slightly relevant.

About 65 percent of the English teachers and 70 percent of the math teachers do not receive district assistance on materials that can be used to prepare students for particular skills to improve test-taking abilities. But of those teachers who do receive this help, most find it relevant or very relevant; about eight to 10 percent of them find it slightly relevant.

Almost 65 percent of the English teachers and about 70 percent of the math teachers indicated that there is no district assistance in teacher interpretation and use of different types of tests and their applications. But once again, most teachers who do receive this assistance note that it is relevant or very relevant to their classroom work; a few indicate that it is at least slightly relevant.

In the matter of tying what they teach to the kinds of skills or content covered on required tests, 60 percent of the English teachers and somewhat more than 70 percent of the math teachers do not receive this kind of help from their districts. Again, those who do, find it mostly relevant or very revelant, with a few finding at least slightly relevant.

Finally, 75 percent of the English teachers and almost 85 percent of the math teachers reported that there is no district training to help teachers use the results of tests to improve their instructional programs. Of the teachers who do receive this training, most find it relevant or very relevant to their classroom work, with a few teachers rating it only as slightly relevant.

With the exception of some district assistance in test administration and test analysis or interpretation, then, the secondary teachers indicate that most of them do not receive the kinds of assistance asked about in the survey; on the other hand, the teachers who do receive assistance in matters related to student assessment, by and large, appear to find it to have specific relevance to their classroom work.

District Training/College Courses

Only 58 of the English teachers and 46 of the math teachers indicated that they have attended district meetings on test selection/construction and/or district testing policies in the past two years. Based on the smaller numbers of teachers responding in the affirmative to a related item in the preceding series, we might suspect that these meetings were more concerned with policy than with test selection/construction. At any rate, of the English teachers attending such meetings, about 40 percent of them indicated that they have attended one to five hours of such meetings; about another 35 percent that they have attended six to 10 hours of these meetings. Of the math teachers responding in the affirmative, about 50 percent have attended one to five hours of such meetings; about another 25 percent have attended six to ten hours of such meetings.

In terms of district inservice on other topics related to student assessment, 63 English teachers and 45 math teachers responded that they have received such inservice. Of the English teachers a little more than 65 percent of them indicated such inservice in the range of one to five hours in the past two years. Of the math teachers, about 65 percent of them noted that their inservice in the last two years amounts to one to five hours; for another 30 percent this inservice amounts to six to 10 hours.

Twenty-eight English teachers and 10 math teachers reported that they have taken college courses in the last two years that were devoted exclusively to student assessment. For 54 percent of the English teachers, the courses they have taken amount to one to five hours; one or two teachers indicate college courses in each of the five-hour intervals between six and 60. Similar patterns hold for the math teachers responding to this item.

District Uses of Assessment Information

The secondary teachers in the sample responded next to four questions dealing with school-administration uses of assessment information in relation to teachers' instructional practices. Again, a little more than 120 English teachers and just under 120 math teachers responded to these questions.

For the first question, dealing with school administration review of test scores with teachers to identify skills or content areas in need of additional emphasis, only about eight percent of the English teachers and 10 percent of the math teachers indicated that this kind of practice happens regularly as part of the school's routine procedures. For about 25 percent of the English teachers but only for about 10 percent of the math teachers this practice happens quite often but not on a regular or routine basis. For a little more than 40 percent of the English teachers and just under 40 percent of the math teachers this happens rarely and on no regular basis. Finally, for about 25 percent of the English teachers and about 38 percent of the math teachers it does not happen at all.

The next question dealt with school administration observation of teaching, reviewing teachers' lesson plans, and/or requiring teachers to write reports to ensure that students' special needs, as shown by test scores, are emphasized. For about 25 percent of the English teachers and about 18 percent of the math teachers this practice is regular and part of the school's routine procedures. For approximately 15 percent of both teacher samples the practice happens quite often but not on a regular or routine basis. For just over 30 percent, again for both teacher samples, the practice is rare and on no regular basis. The practice does not happen at all for 30 percent of the English teachers and about 37 percent of the math teachers.

Being required to turn in the scores/grades of tests or assignments that they routinely give in class appears to be a regular and routine procedure for about seven percent of the English teachers and five percent of the math teachers; these approximate percentages hold for quite frequent occurrence which is neither routine nor regular, and for rare occurrence on no regular basis. In each case, the percentages are slightly higher for the English teachers. The practice does not happen at all for about 73 percent of the English teachers and almost 80 percent of the math teachers.

The last question in this series asked the teachers if their school administration evaluates their teaching on the basis of student test scores and/or establishes specific test-score grades for the students and the teacher to meet. This practice is regular and routine for only one percent of the English teachers and about three percent of the math teachers. It happens quite often but not on a regular or routine basis for about seven percent of the English teachers and three percent of the math teachers. It happens rarely and on no regular basis for seven or eight percent of each teacher sample, and does not happen at all for the approximately 85 percent remaining for each sample.

District Reporting of Test Results

The final background question in the survey asked questions on test turn-around time, the usefulness of test reporting formats, and whether the districts encourage teachers to emphasize the basic skills. Response rates were again around 120 for each teacher sample.

Approximately 32 English teachers (26%) and 28 math teachers (24%)

indicated that their district returns test results quickly enough that teachers can use them for modifying their instruction. About 43 English teachers (36%) and 32 math teachers (28%) noted that the results are returned too slowly for the teacher to use them in modifying teaching. Ten or 11 teachers (9%) in each sample responded that the district does not return their students' test results, and about 35 English teachers (30%) and 45 math teachers (39%) indicated that the question does not apply.

In response to whether the district reports back students' test results in a way that facilitates teachers' use of the information, 56 English teachers (46%) and 40 math teachers (35%) indicated that detailed results are provided in a useful format. This finding appears to be a little at odds with some of the responses to the items immediately above. It may be that while some teachers receive results too late to modify instruction, they do make other uses of the information. About 31 English teachers (25%) and 28 math teachers (24%) responded that the district provides little useful information in the way of test results, and 35 English teachers (29%) and 47 math teachers (41%) that the question does not apply.

The last question in this series asked whether the district has encouraged teachers, in connection with an assessment program, to emphasize the teaching of basic skills. About 107 English teachers (88%) indicated that their districts do follow this practice, while the remaining 12% that their districts do not. For the math teachers approximately the same percentages hold.

Teacher Attitude Toward Testing and Test-Related Issues

A number of items on the survey probed teachers' attitudes toward testing and test-related matters. English teachers' response rates to these items ran from 103 to 122; for the math teachers, response rates were from 97 to 115. Table 20 below shows the percentages of English teachers and math teachers who strongly agreed or agreed with a series of statements on the topic of concern.

Table 20:

Teacher Attitude Toward Testing and Test-Related Issues

Item	Percentage of Teachers In Agreement	
	English	Math
Testing motivates my students to study harder.	82	92
Commercial tests are usually of high quality.	47	43
The content (or skills) on most required tests is very similar to the content (or skills) that I teach.	72	76
The pressure that testing exerts on the schools has a generally beneficial effect.	59	72
Recently, I have been spending more teaching time preparing my students to take required tests.	44	34
The tests developed in our district are very good.	59	57
The curriculum today demands more complex student thinking than in the past.	66	51
Teachers should not be held accountable for students' scores on standardized achievement tests or tests of minimum competency.	60	67
In our school, students are more rigidly tracked than they were two or three years ago.	40	31
Tests of minimum competency/proficiency/functional literacy should be required of all students for promotion at certain grade levels or for high school graduation.	90	92
Tests of minimum competency are frequently unfair to particular students.	49	30
As a result of minimum competency tests (and similar programs), parents are contacting schools about their children more frequently or in greater numbers.	39	40

Table 20:
(continued)

Item	Percentage of Teachers in Agreement English	 Math
Tests of minimum competency have affected (would affect) the amount of time I can spend teaching subjects or skills that the tests do not cover.	65	42
In our school, testing programs are generally held to be much less important than the social problems with which we are concerned.	33	44
Basic skills teaching (including remedial work) is now consuming a substantially increased proportion of our school's educational resources.	85	78
The proportion of school's resources now allocated to basic skills teaching is so great as to detract from the quality of our total educational program.	33	24

Some fairly clear trends emerge from these data. On the one hand, the vast majority of secondary teachers from both samples state agreement with the use of minimum competency tests for promotion or graduation. On the other hand, while most math teachers do not believe that these tests are unfair to certain kinds of students, about 30 percent of them do, and about 50 percent of the English teachers would agree.

The great majority of both samples agree that testing motivates their students to study harder, yet about 60 percent of the English teachers and 70 percent of the math teachers feel that teachers should not be held accountable for students' scores on standardized or minimum competency tests. On the other hand, sizable numbers of teachers in both samples disagree, and believe that teachers should be held accountable for student performance on these tests.

At the crux of this issue, perhaps, is the kind of test used, its purpose, and its origin. For example, the majority of both samples appear dubious about the quality of commercial tests; greater numbers of teachers in both samples appear rather more comfortable with the tests developed in their own districts. Perhaps teachers accept being held accountable for students' scores on locally developed and locally "normed" tests. This supposition might be borne out by the high levels of teacher agreement that the content or skills on most required tests is similar to the content they teach, especially should these required tests be locally developed and driven by the local curriculum or come with the curriculum accepted and used by the teachers.

The great majority of both samples agree that basic skills teaching is now consuming an increasing proportion of their schools' educational resources, yet do not appear to believe that this allocation is so great as to detract from the quality of their schools' total education program. On the other hand, while teachers seem to support the need to teach in the basic skills, some of them are more reserved about the curricular effect of minimum competency testing. For example, most English teachers agree that tests of minimum competency affect the amount of time they can spend teaching content/skills that these tests do not cover; this may suggest overemphasis of reading comprehension testing to the detriment of other skills held important by English teachers. On the other hand, almost 60 percent of the math teachers do not agree that tests of minimum competency affect the amount of time they can spend teaching content/skills not covered by the tests. Perhaps math teachers take a different view of subject breadth versus basis skills; perhaps they find a better fit between required tests and content taught.

A somewhat varied picture of tests and testing seems to exist, starting with the majority agreement that testing programs are generally held to be more important than the social programs teachers are concerned with. Yet the majority of both samples agree that testing has a generally beneficial effect; they also agree that their schools are not spending increasing instructional time to prepare students to take required tests, and that their students are not becoming more rigidly tracked. Given the finding that most secondary teachers believe that the curriculum today demands more complex student thinking than in the past, teacher perception about tracking is important. Rigid tracking, especially if done on the basis of tests not seen as accurate by teachers, might be seen as affecting their potential to stimulate students.

Teacher Uses of Assessment Results

The secondary teachers responded to a set of questions on how they use various kinds of information in their decision making about students. The decision concerns they responded to were the same as those queried in the elementary sample -- (1) planning teaching at the start of the school year; (2) initial grouping or placement of students; (3) changing a student from one group or curriculum to another; and (4) assigning students' report card grades. The data in the table following indicate the percentage of teachers who rated a given information source as crucial or important for the decision purpose. Numbers in parenthesis reflect percentages of teachers reporting that the assessment information is not available. Response rates continue to be in the range of 115 to 120 for each sample.

Table 21:

Teacher Use of Assessment Information for Different Decision-Making Purposes

(Percentages reporting use of this information for the specified purpose)

	Planning Teaching at Beginning of School Year		Initial Grouping of Students		Changing a Student from One Group or Curriculum to Another		Deciding on Students' Report Card Grades	
Source/Kind of Information	English	Math	English	Math	English	Math	English	Math
Previous teacher's comments, reports, grades	32 (8)	29 (9)	38 (8)	38 (9)				
Students' standardized test scores	51 (2)	25 (3)	53 (3)	29 (10)	65 (1)	40 (11)	15* (14)	7* (22)
Students' scores on district continuum or minimum competency tests	50 (19)	27 (18)	46 (20)	36 (23)	52 (19)	40 (23)	9 (27)	3* (27)
My previous teaching experience	98 --	96 --						
Results of tests included with curriculum being used			41 (29)	34 (36)	58 (13)	40 (25)	45 (17)	30 (26)
Results of other special placement tests			42 (26)	28 (34)				
Results of special tests developed or chosen by my school					53 (23)	32 (39)	32 (25)	26 (33)
Results of tests I make up			87 --	72 (9)	92 --	92 (3)	100 --	98 --
My own observations and students' classroom work			97 --	85 --	99 --	96 --	99 --	97 --

*These ratings are for "important" only; they do not reflect any "crucial" ratings.

As was the case with the elementary school teachers, individual teachers' previous experience is by far the most important source of information for most teachers as they plan instruction at the beginning of the school year. For the English teachers, students' scores on standardized tests and their scores on district continua or tests of minimum competency are held as important by about half of the sample, followed by previous teachers' comments with about 30 percent. In addition, for teachers' comments and standardized and district continua/minimum competency tests, another 20 to 30 percent of the English teachers find them to be slightly important in this decision area. Note that for students' scores on district continua/minimum competency tests, almost 20 percent of the English teachers report this kind of assessment information is not available to them.

These patterns are of the same order as those obtaining for the math teachers, with the exception that only about 25 percent of these raters find standardized and district continua/minimum competency test scores to be crucial or important. For teacher comments, another 40 percent of the math teachers find them to be slightly important. Again, a sizable number of math teachers (18%) indicate that district continua/ minimum competency test data are not available to them.

In making their decisions about initial grouping or placement of students, teachers' own observations and the results of tests they make up themselves are deemed most important by most of the English and math teachers. Previous teachers' comments are the same for both populations with almost 40 percent finding them crucial or important, and another 25 percent finding them to be slightly important.

Again, as was the case with the elementary teachers, note that students' scores on "formal" tests continue to have importance for some teachers as they make their initial grouping decisions; this trend is somewhat more pronounced for the English teachers, especially in the case of standardized test scores. These more formal measures, further, are slightly important for anywhere from 15 to 30 percent of the teachers depending on the particular source of information. Note once again that for a sizable number of teachers, certain kinds of test information are reported as not being available: 10 percent of the math teachers make this statement for standardized tests; about 20 percent each for English and math report there are no scores on district continua/minimum competency tests; depending on the particular measure being cited, anywhere from 25 to 35 percent of the teachers state there is no information available from tests that are part of their curricula or from other special placement tests. While non-availability of some of these measures (e.g., standardized tests, curriculum tests) is not too surprising early in the year when initial grouping decisions are being made, the unavailability of other special placements tests for a fair number of teachers may be noteworthy.

The picture with regard to teachers' decisions about changing a student from one group or curriculum to another looks quite balanced. Once again, teacher observations and results of their own tests are the most important sources of information for most teachers. But note that both samples demonstrate that there is still some reported importance for standardized tests in this decision area. Particularly for the English teachers, standardized tests, albeit in conjunction with other kinds of assessment information, are still important in decisions being made once the school year is well underway. Similar patterns hold for district continua/minimum competency tests, tests that are part of curricular materials, and results of special tests developed or chosen by the school. And again, a fair number of teachers also report these devices to be slightly important in their decision making.

While some of the findings reflecting the unavailability of certain kinds of assessment information early in the school year are not surprising, it is a little more surprising that so many teachers report their non-availability once the school year is underway and decisions about instructional and classroom management modifications are being made: in this regard, about 10 percent of the math teachers report that no standardized test data are available; roughly 20 percent of each sample report that information from district continua or minimum competency tests is not available to them; almost 15 percent of the English teachers and 25 percent of the math teachers report non-availability of information from curriculum tests; almost one quarter of the English teachers and about 40 percent of the math teachers report the same for special tests developed or chosen by the school.

With regard to making decisions about students' report card grades, results of their own tests and other observations of students remain of greatest importance for most teachers. Results of curriculum tests appear next in order of importance as reflected by percentages of teachers, followed by results of tests developed or chosen by their school.

Note that the indices of non-availability of information from a given measure remain fairly constant between decisions involving student changes and decisions about their report card grades. That is, where information is reported as unavailable for teacher decisions during the school year or semester, it also appears to be equally unavailable at or near the end of the year/semester. Perhaps for some teachers these measures simply do not exist; for others it may be that the results of certain measures are not made available to teachers when they are needed for a given decision; perhaps for some tests the results are administered and filed centrally and are never provided to teachers. The latter two cases might be distinct possibilities based on teachers' responses earlier in this section on the manner in which test results are returned to them.

REFERENCES

Airasian, P.W. The effects of standardized testing and test information on teachers' perceptions and practices. Paper presented at the annual meeting of the American Educational Research Association. San Francisco, California, 1979.

Angel, J.L. National, state, and other external testing programs. Review of Educational Research, 1968, 38(1), 85-91.

Baker, E.L. Achievement testing in urban schools: New numbers. Paper presented at CEMREL Conference on Urban Education, 1978. (Also CEMREL Monograph on Urban Education, St. Louis, Missouri: CEMREL, 1980.)

Bank, A.,& Williams, R. Evaluation in school districts: Organizational perspectives. CSE Monograph Number 10. Los Angeles: Center for the Study of Evaluation, 1981, in press.

Borko, H. An examination of some factors contributing to teachers' preinstructional classroom organization and management decisions. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada, 1978.

Boyd, J., McKenna, B.H., Stake, R.E., & Yashinski, J. A study of testing practices in the Royal Oak (Michigan) public schools. Royal Oak, Michigan: Royal Oak City School District, 1975. (ERIC Document Reproduction Service No. ED 117 161.)

Carducci-Bolchazy, M. A survey of the use of readiness tests. Reading Horizons, 1968, 18(3), 209-212.

Crew, J.L., & White, E.N. Criterion Referenced testing: Usages in some member systems of the Council of Great City Schools. Journal of Negro Education, 1978, 47(2), 159-167.

Ebel, R.L. Improving the competence of teachers in educational measurement. In J. Flynn & H. Garber (Eds.), Assessing behavior: Readings in educational and psychological measurement. Reading, Massachusetts: Addison-Wesley, 1967.

Gorth, W.P., & Perkins, M.R. A study of minimum competency testing programs: A final report. Washington, D.C.: Office of Testing, Assessment, and Evaluation, National Institute of Education, 1980.

Goslin, D.A. The use of standardized ability tests in American secondary schools and their impact on students, teachers, and administrators. New York: Russell Sage Foundation, 1965.

Goslin, D.A., Epstein, R., & Hallock, B.A. The use of standardized tests in elementary schools. Second Technical Report. New York: Russell Sage Foundation, 1965.

Goslin, D.A. Teachers and testing. New York: Russell Sage Foundation, 1967.

Hotvedt, M.O. Teacher use of standardized tests. Health Service Center at Houston, University of Texas, Houston, Texas, 1978.

Infantino, R. Results of the NYSEC survey on testing. The English Record, 1975, 26(2), 4-10.

Jessen, R.J. Probability sampling with marginal constraints. Journal of the American Statistical Association, 1970, 65(330), 776-796.

Kaufman, J.D. State assessment programs: Current status and a look ahead. Denver, Colorado: Scholastic Testing Service, Inc., 1979.

Kirkland, M. The effects of tests on students and schools. Review of Educational Research, 1971, 41(4), 303-350.

Kitsuse, J.I., & Cicourel, A.V. The educational decisionmakers. Indianapolis, Indiana: Bobbs-Merrill, 1963.

Leiter, K.C.W. Ad hocing in the schools: A study of placement practices in two kindergartens. In A.V. Cicourel, (Ed.), Language use and school performance. New York: Academic Press, 1974.

Lortie, D.C. Schoolteacher: A sociological study. Chicago: The University of Chicago Press, 1975.

Lyon, C.D., Doscher, L., McGranahan, P., & Williams, R. Evaluation and school districts. Los Angeles: Center for the Study of Evaluation, 1978.

Mehan, H. Ethnomethodology and education. *Proceedings of the Second Annual Conference of the Sociology of Education Association*, 1974.

National Evaluation Systems. Proposal submitted to the National Institute of Education for the review and evaluation of minimum competency testing programs, 1978.

Resnick, L., & Resnick, D. *The social functions of educational testing. A proposal submitted to the Carnegie Corporation of New York, 1978.*

Rudman, H.C. The superintendent and testing: Implications for the curriculum. *Measurement in Education*, 1975, *8*(4), 1-8.

Rudman, H.C., Kelly, J.L., Wanous, D.S., Mehrens, W.A., Clark, C.H., & Porter, A.C. *Integrating testing with instruction. A review 1922-1980*. East Lansing, Michigan: Institute for Research on Teaching, 1980.

Salmon-Cox, L. *Teachers and tests: What's really happening?* Paper presented at the annual meeting of the American Educational Research Association, Boston, Massachusetts, 1980. (Also *Phi Delta Kappan*, 1981, 631-634.)

Schutz, A. *Collected Papers I: The problem of social reality*. M. Natanson (Ed.) The Hague, Netherlands: Martinus Nijhoff, 1962.

Shavelson, R.J., Russo, N.A., & Borko, H. Experiments on some factors contributing to teachers' pedagogical decisions. *Cambridge Journal of Education*, 1977, *7*, 51-70.

Sproull, L., & Zubrow, D. *Test information in the suburban school district*. Paper presented at the annual meeting of the American Educational Research Association, Boston, Massachusetts, 1980.

Stetz, F., & Beck, M. *Teachers' opinions of standardized test use and usefulness*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, California, 1979.

Yeh, J. *Test use in schools*. Los Angeles: Center for the Study of Evaluation, University of California, Los Angeles. Work Unit 4, Studies in Measurement and Methodology, June, 1978.